PROBABILIDAD Y ESTADÍSTICA
Aplicaciones y métodos

George C. Canavos
VIRGINIA COMMONWEALTH UNIVERSITY

Traducción:
Edmundo Gerardo Urbina Medal
Departamento de Ingeniería Eléctrica
UAM Ixtapalapa

Revisión Técnica:
Gustavo Javier Valencia Ramírez
Doctor en Matemáticas
Profesor Titular
Departamento de Matemáticas
Facultad de Ciencias
UNAM



MÉXICO • BUENOS AIRES • CARACAS • GUATEMALA
LISBOA • MADRID • NUEVA YORK • PANAMÁ • SAN JUAN
SANTAFÉ DE BOGOTÁ • SANTIAGO • SÃO PAULO
AUCKLAND • HAMBURGO • LONDRES • MILÁN • MONTREAL
NUEVA DELHI • PARÍS • SAN FRANCISCO • SINGAPUR
ST. LOUIS • SIDNEY • TOKIO • TORONTO

PROBABILIDAD Y ESTADÍSTICA
Aplicaciones y métodos

McGRAW-HILL/INTERAMERICANA DE MEXICO, S.A. DE C.V.
Atlacomulco 499-501, Fracc. Industrial San Andrés Atoto
53500 Naucalpan de Juárez, Edo. de México
Miembro de la Cámara Nacional de la Industria Editorial, Reg. Núm. 1890

ISBN 968-451-856-0

Traducido de la primera edición en inglés de
APPLIED PROBABILITY AND STATISTICAL METHODS

ISBN 0-316-12778-7

1203456789 P.E.-87 9076543218

Impreso en México Printed in Mexico

Esta Obra se terminó de imprimir en Enero de 1998 en Programas Educativos, S.A. de C.V. Calz. Chabacano No. 65-A Col. Asturias Delegación Cuauhtémoc C.P. 06850 México, D.F. Empresa Certificada por el Instituto Mexicano de Normalización y Certificación A.C. bajo la Norma ISO-9002: 1994/NMX-CC-004: 1995 con el Núm. de Registro RSC-048

Se tiraron 2500 ejemplares

A mi madre,
y a Athena, Alexis y Costa

Contenido

CAPÍTULO TRES

Variables aleatorias y distribuciones de probabilidad 52

CAPÍTULO CUATRO

Algunas distribuciones discretas de probabilidad 88

CAPÍTULO CINCO

Algunas distribuciones continuas de probabilidad 130

CAPÍTULO OCHO

Estimación puntual y por intervalo 251

CAPÍTULO NUEVE

Prueba de hipótesis estadísticas 303

CAPÍTULO DIEZ

Pruebas de bondad de ajuste y análisis de tablas de contingencia 362

CAPÍTULO ONCE

Métodos para el control de calidad y muestreo para aceptación 379

CAPÍTULO DOCE

Diseño y análisis de experimentos estadísticos 401

CAPÍTULO TRECE

Análisis de regresión: el modelo lineal simple 443

CAPÍTULO CATORCE

Análisis de regresión: el modelo lineal general

CAPITULO QUINCE

Métodos no paramétricos 572

Prefacio

Este libro se planeó como una introducción a la teoría de la probabilidad y a la inferencia estadística, para toda persona interesada en las disciplinas aplicadas; economía y finanzas, ingeniería y ciencias físicas y de la vida. No es necesario ningún conocimiento previo de probabilidad y estadística, aunque se espera que el lector se encuentre familiarizado con los fundamentos del cálculo diferencial e integral. El libro hace hincapié en las aplicaciones. El rigor matemático se emplea únicamente con el fin de exponer las bases de la probabilidad y de la estadística, lo que, en opinión del autor, es un ingrediente necesario para la aplicación efectiva de los métodos. El texto intenta proporcionar al estudiante un conocimiento que vaya más allá de lo superficial, sin abrumarlo con teoría excesiva. En este sentido, la obra brinda la oportunidad de reforzar el "porqué", además de presentarle el "cómo" de la aplicación.

A través del texto, cada concepto o método se ilustra con ejemplos reales que se expresan de manera que el lector pueda obtener una comprensión intuitiva del concepto. La mayor parte del desarrollo de la inferencia estadística se fundamenta en el punto de vista de la teoría del muestreo. También se explora el enfoque bayesiano para dar la perspectiva adecuada. Asimismo, se estudian las suposiciones de los métodos estadísticos y se dan respuestas a preguntas del tipo "qué pasa si..." Además, en muchos ejemplos se emplearon paquetes de programas para computadora y técnicas de simulación, con el propósito de ilustrar y reforzar los puntos presentados.

El material que abarca el libro demuestra ser suficiente para realizar un curso de dos semestres sobre probabilidad y métodos estadísticos. Por otra parte, es posible reordenar el material y así ofrecer variedad de cursos, como un curso de un semestre sobre distribuciones de probabilidad y sus aplicaciones, en el que se empleen los capítulos 1 a 7; un curso de dos trimestres sobre los fundamentos de la probabilidad y de los métodos estadísticos, con los capítulos 1 a 10; o un curso en análisis de varianza y métodos de regresión, con los capítulos 9, 12, 13 y 14. El alcance de los temas que se tratan es amplio, extenso y proporcionan al profesor la oportunidad de recalcar ciertos temas u omitir otros. Que el libro pueda emplearse a nivel licenciatura o a nivel de graduados, depende tanto de las necesidades particulares como de los conocimientos previos de los lectores.

Después de un análisis razonablemente completo sobre la estadística descriptiva (Cap. 1), el libro está dividido en probabilidad (Caps. 2-7) y métodos esta-

dísticos (Caps. 8-15). En los capítulos 2 y 3 se presentan los conceptos básicos de probabilidad, variable aleatoria y distribución de probabilidad. Los capítulos 4 y 5 contienen una exposición bastante completa de las distribuciones de probabilidad discretas y continuas, así como sus aplicaciones. En estos capítulos se investigan, comparan y contrastan propiedades de distribuciones como la binomial, de Poisson, normal, beta, gama y de Weibull, entre otras, proporcionando áreas de aplicación para cada una. Dado el creciente papel de las computadoras y las técnicas de simulación, se dedica una sección del capítulo 5 a la valoración de varios métodos de generación de valores aleatorios, en cada una de las distribuciones estudiadas. En el capítulo 6 se exponen las distribuciones de probabilidad conjunta y condicional. En este contexto, se introducen los conceptos de distribuciones *a priori* y *a posteriori*, para el punto de vista bayesiano.

El capítulo siete funciona como transición entre la probabilidad y la inferencia estadística. En éste se plantean los importantes conceptos de muestra aleatoria y distribución de muestreo. En el capítulo 8 se presentan los métodos de estimación, tanto puntual como de intervalo. También se estudian los límites de tolerancia independientes de la distribución y aquéllos cuyo fundamento es la distribución normal. En el capítulo 9 se exploran las bases de la inferencia estadística y se presentan las pruebas de hipótesis para medias, varianzas y proporciones. El capítulo 10 detalla el uso de la distribución chi-cuadrada, tanto para determinar la bondad del ajuste, como para tablas de contingencia, mientras que el capítulo 11 introduce al lector en los conceptos básicos del control de calidad estadístico y a los procedimientos para aceptar una muestra. En el capítulo 12 se presentan el diseño de experimentos estadísticos y el análisis de varianza, tanto para experimentos de un solo factor como para dos. En los capítulos 13 y 14 se trata, de manera prolija, el análisis de regresión; además, se examinan con detalle temas como: errores autocorrelacionados, análisis de residuos, mínimos cuadrados con factores de peso, multicolinealidad y distintas formas para determinar el mejor conjunto de variables de predicción. Al concluir, el capítulo 15 explora y compara algunos de los procedimientos no paramétricos más útiles.

Al final del capítulo 1 y del 13 se encuentra un apéndice en que se revisa la notación sumatoria y del álgebra matricial. Las demostraciones de los teoremas más importantes se encuentran, para los lectores cuyas inclinaciones son más hacia la teoría, en los apéndices de los capítulos 4, 5 y 7. En el apéndice del libro se encuentran once tablas estadísticas. Se intentó, hasta donde fue posible, uniformar la estructura de éstas; por ejemplo, se encuentran tabulados valores para las distribuciones binomial, de Poisson, hipergeométrica y normal, además de los valores cuantiles para las distribuciones chi-cuadrada, t de Student y F. Las tablas para las distribuciones anteriores, excepto la hipergeométrica, se generaron mediante algunas subrutinas del paquete IMSL (*International Mathematical and Statistical Libraries*). La similitud con las tablas estadísticas, ya establecidas, es excelente. Los paquetes para computadora Minitab y SAS (*Statistical Analysis System*) se emplearon con objeto de ilustrar las técnicas del análisis de regresión (Caps. 13 y 14). Se supone que el lector tiene acceso a algunos de estos paquetes o a otros similares, como el SPSS (*Statistical Package for the Social Sciences*) y BMDP (*Biomedical Programs*).

Deseo agradecer a todas las personas que por muchos años, y de una forma y otra, desempeñaron un papel directo o indirecto para que este libro fuese posible; en particular, al Departamento de Estadística del Virginia Polytechnic Institute y de la State University, donde aprendí estadística por primera vez; al NASA's Langley Research Center, donde se me dio la oportunidad de continuar mis estudios de estadística, y a la Virginia Commonwealth University, donde generalmente enseño estadística. También deseo agradecer la ayuda de John Koutrouvelis, del Departamento de Ciencias Matemáticas de la Virginia Commonwealth University, pues con sus críticas contribuyó de manera significativa en los capítulos sobre probabilidad. Además, extiendo mi gratitud a las siguientes personas, quienes me proporcionaron sugerencias muy útiles durante todas las etapas del desarrollo del manuscrito: Arlene S. Ash, de la Boston University; Bruce K. Blaylock, del Virginia Polytechnic Institute y de la State University; George W. Brown, de la University of California, en Irvine; Donald R. Burleson, del Rivier College; John M. Burt, de la University of New Hampshire; Dean H. Fearn, de la California State University en Hayward; Richard H. Lavoie, del Providence College; Stephen Meeks, de la Boston University; Chester Piascik, del Bryant College; Ramona L. Trader, de la University of Maryland, y George D. Weiner, de la Cleveland State University.

Extiendo también mi aprecio a Carolyn England, K.W. Hall y Jamie Stokes, quienes compartieron la labor de escribir todas las versiones del manuscrito. Gracias, de manera especial, al grupo editorial de Little, Brown and Company, y en particular a Elizabeth Schaaf por su valiosa ayuda. Por último deseo agradecer a mi familia su paciencia, comprensión y aliento durante el tiempo en que escribí el libro.

George C. Canavos

CAPÍTULO UNO

Introducción y estadística descriptiva

1.1 Introducción

Para mucha gente, *estadística* significa descripciones numéricas. Esto puede verificarse fácilmente al escuchar, un domingo cualquiera, a un comentarista de televisión narrar un juego de fútbol. Sin embargo, en términos más precisos, la estadística es el estudio de los fenómenos aleatorios. En este sentido la ciencia de la estadística tiene, virtualmente, un alcance ilimitado de aplicaciones en un espectro tan amplio de disciplinas que van desde las ciencias y la ingeniería hasta las leyes y la medicina. El aspecto más importante de la estadística es la obtención de conclusiones basadas en los datos experimentales. Este proceso se conoce como *inferencia estadística*. Si una conclusión dada pertenece a un indicador económico importante o a una posible concentración peligrosa de cierto contaminante, o bien, si se pretende establecer una relación entre la incidencia de cáncer pulmonar y el fumar, es muy común que la conclusión esté basada en la inferencia estadística.

Para comprender la naturaleza de la inferencia estadística, es necesario entender las nociones de *población* y *muestra*. La población es la colección de toda la posible información que caracteriza a un fenómeno. En estadística, población es un concepto mucho más general del que tiene la acepción común de esta palabra. En este sentido, una población es cualquier colección ya sea de un número finito de mediciones o una colección grande, virtualmente infinita, de datos acerca de algo de interés. Por otro lado, la muestra es un subconjunto representativo seleccionado de una población. La palabra *representativo* es la clave de esta idea. Una buena muestra es aquella que refleja las características esenciales de la población de la cual se obtuvo. En estadística, el objetivo de las técnicas de muestreo es asegurar que cada observación en la población tiene una oportunidad igual e independiente de ser incluida en la muestra. Tales procesos de muestreo conducen a una *muestra aleatoria*. Las observaciones de la muestra aleatoria se usan para calcular ciertas características de la muestra denominadas *estadísticas*. Las estadísticas se usan como base para hacer inferencias acerca de ciertas características de la población, que reciben el nombre de

parámetros. Así, muchas veces se analiza la información que contiene una muestra aleatoria con el propósito principal de hacer inferencias sobre la naturaleza de la población de la cual se obtuvo la muestra.

En estadística la inferencia es inductiva porque se proyecta de lo específico (muestra) hacia lo general (población). En un procedimiento de esta naturaleza siempre existe la posibilidad de error. Nunca podrá tenerse el 100% de seguridad sobre una proposición que se base en la inferencia estadística. Sin embargo, lo que hace que la estadística sea una ciencia (separándola del arte de adivinar la fortuna) es que, unida a cualquier proposición, existe una medida de la confiabilidad de ésta. En estadística la confiabilidad se mide en términos de probabilidad. En otras palabras, para cada inferencia estadística se identifica la probabilidad de que la inferencia sea correcta.

Los problemas estadísticos se caracterizan por los siguientes cuatro elementos:

1. La población de interés y el procedimiento científico que se empleó para muestrear la población.
2. La muestra y el análisis matemático de su información.
3. Las inferencias estadísticas que resulten del análisis de la muestra.
4. La probabilidad de que las inferencias sean correctas.

El enfoque precedente para la inferencia estadística descansa únicamente en la evidencia muestral. Éste es denominado *teoría del muestreo* o enfoque *clásico* de la inferencia estadística y para la mayor parte de ésta, será el que se tome en este libro. Sin embargo, también se tratará de incorporar ocasionalmente otro punto de vista conocido como *inferencia bayesiana.* Esta forma de abordar la inferencia estadística utiliza la combinación de la evidencia muestral con otra información, generalmente proporcionada por el investigador del problema. Tal información descansa de manera fundamental en la convicción o grado de creencia del investigador con respecto a las incertidumbres del problema, antes de que se encuentre disponible la evidencia muestral. Este grado de creencia puede basarse en consideraciones como los resultados conocidos, que son producto de investigaciones previas. Es importante que el lector comprenda que el objetivo de los procedimientos clásico y bayesiano descansa en la evaluación de las incertidumbres basadas en la probabilidad.

Para comprender la esencia del muestreo aleatorio y de la inferencia estadística, es necesario entender como primer punto, la naturaleza de una población en el contexto de la probabilidad y de los modelos probabilísticos. Estos temas se examinan con detalle en los capítulos dos a seis.

Este capítulo tratará brevemente las *estadísticas descriptivas.* A pesar de que éstas son sencillas desde el punto de vista matemático, son valiosas en casos donde se encuentra disponible la población completa y no existe incertidumbre, o cuando se tienen a la mano grandes conjuntos de datos que pueden o no considerarse como muestras aleatorias. Si un conjunto grande se considera como muestra aleatoria de una población, la estadística descriptiva puede ir tan lejos como la distribución general de valores, al dar una evidencia empírica y otras características de la población. Esta evidencia tiene un apreciable valor puesto que afirma ciertas suposiciones que deben formularse en la aplicación de la inferencia estadística.

1.2 Descripción gráfica de los datos

Una descripción informativa de cualquier conjunto de datos está dada por la frecuencia de repetición u arreglo distribucional de las observaciones en el conjunto. Para apreciar lo necesario de un resumen de datos, considere el ejemplo del Servicio de Hacienda Interno (SHI) que se encarga de recibir y procesar millones de declaraciones de ingresos durante todo el año. Es dudoso que el SHI pueda descubrir los patrones ocultos de ingresos e impuestos examinando simplemente la información contenida en las declaraciones. Similarmente, el Departamento del Censo no podría avanzar mucho al analizar los datos del censo, si éstos no pudiesen visualizarse. Para identificar los patrones en un conjunto de datos es necesario agrupar las observaciones en un número relativamente pequeño de clases que no se superpongan entre sí, de tal manera que no exista ninguna ambigüedad con respecto a la clase a que pertenece una observación en particular. El número de observaciones en una clase recibe el nombre de *frecuencia de clase,* mientras que el cociente de una frecuencia de clase con respecto al número combinado de observaciones en todas las clases se conoce como la *frecuencia relativa* de esa clase. Las fronteras de la clase se denominan límites, y el promedio aritmético entre los límites superior e inferior recibe el nombre de *punto medio* de la clase. Al graficarse las frecuencia relativas de las clases contra sus respectivos intervalos en forma de rectángulos, se produce lo que comúnmente se conoce como *histograma de frecuencia relativa* o *distribución de frecuencia relativa*. Esta última es la que puede hacer evidentes los patrones existentes en un conjunto de datos.

Como ilustración, los datos de la tabla 1.1 representan las frecuencias de unidades vendidas por día de un determinado producto por una compañía. El histograma de frecuencia relativa se construye graficando en el eje vertical la frecuencia relativa y en el eje horizontal las fronteras inferiores de cada clase, como se ilustra en la figura 1.1.

El número de clases que se emplea para clasificar los datos en un conjunto depende del total de observaciones en éste. Si el número de observaciones es relativamente pequeño, el número de clases a emplear será cercano a cinco, pero general-

TABLA 1.1 Frecuencias para el número de unidades vendidas de cierto producto

Número de unidades vendidas (Clase)	*Frecuencia de la clase*	*Frecuencia relativa*
80–89	7	7/100 = 0.07
90–99	20	20/100 = 0.20
100–109	5	5/100 = 0.05
110–119	11	11/100 = 0.11
120–129	11	11/100 = 0.11
130–139	12	12/100 = 0.12
140–149	6	6/100 = 0.06
150–159	23	23/100 = 0.23
160–169	5	5/100 = 0.05
Total	100	1.00

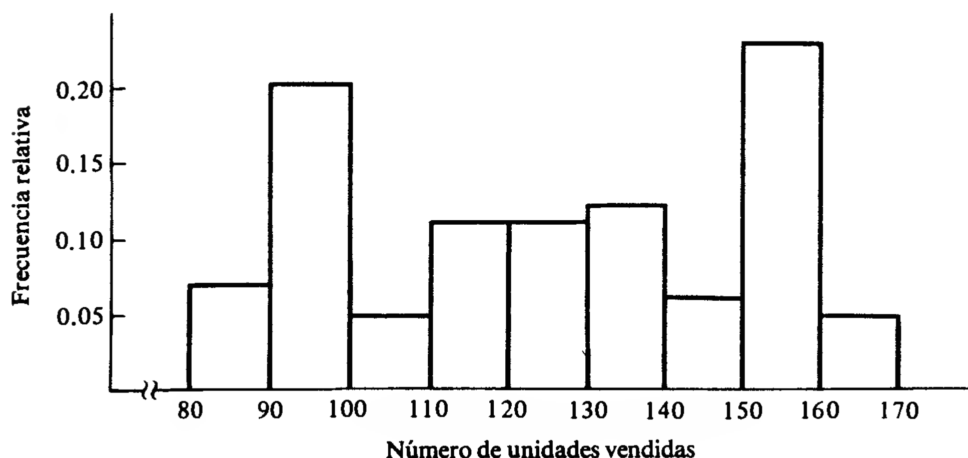


FIGURA 1.1. Histograma de frecuencia relativa para el número de unidades vendidas

mente nunca menor que este valor. Si existe una cantidad sustancial de datos, el número de clases debe encontrarse entre ocho y doce y generalmente no existirán más de 15 clases. Un número muy pequeño de clases puede ocultar la distribución real del conjunto de datos, mientras que un número muy grande puede dejar sin observaciones a algunas de las clases, limitando de esta forma su uso. A manera de ilustración, si se reducen las nueve clases a sólo tres, en el ejemplo anterior, como se indica en la tabla 1.2, el histograma de frecuencia relativa resultante (Fig. 2) es muy diferente al mostrado en la figura 1.1.

Una buena práctica es la creación de clases que tengan una longitud igual. Esto puede lograrse tomando la diferencia entre los dos valores extremos del conjunto de datos y dividiéndola entre el número de clases; el resultado será aproximadamente la longitud del intervalo para cada clase. Sin embargo, existen casos donde esta regla no puede o no debe aplicarse. Por ejemplo, si se tuviera a la mano la lista de impuestos de SHI pagados por la población en un año, estas cantidades pueden encontrarse

TABLA 1.2 Frecuencia para el número de unidades vendidas de cierto producto

Número de unidades vendidas (Clase)	*Frecuencia de la clase*	*Frecuencia relativa*
80–109	32	32/100 = 0.32
110–139	34	34/100 = 0.34
140–169	34	34/100 = 0.34
Total	100	1.00

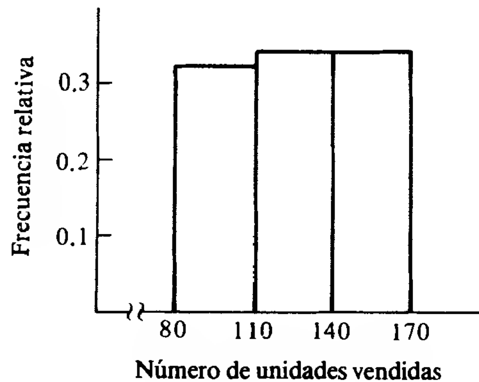


FIGURA 1.2 Histograma modificado para el número de unidades vendidas

en un intervalo de \$0 a \$1 000 000. Aun a pesar de que se eligiesen 20 clases para la distribución de frecuencia relativa, con intervalos de igual longitud, cada clase tendría una cobertura de \$50 000. Lo anterior daría origen a una situación en la que casi todas las observaciones caerían en la primera clase. Para casos como éste es preferible seleccionar una escala más pequeña en el extremo inicial que la utilizada para el extremo superior. Esta elección aclarará el patrón de la distribución.

Los siguientes ejemplos ilustran estos conceptos.

Ejemplo 1.1 De acuerdo con la revista *Informes al Consumidor* en su número de febrero de 1980, las cuotas anuales de 40 compañías para un seguro de \$25 000 para hombre de 35 años de edad son las siguientes:

\$ 82	85	86	87	87	89	89	90	91	91
92	93	94	95	95	95	95	95	97	98
99	99	100	100	101	101	103	103	103	104
105	105	106	107	107	107	109	110	110	111

Establecer un esquema de agrupamiento para este conjunto de datos y determinar las frecuencias relativas.

Dado que la diferencia entre los dos valores extremos del conjunto es de sólo \$29, puede ser razonable agrupar los datos en clases con intervalos de igual longitud. Supóngase que se decide utilizar seis clases; entonces el intervalo de cada clase será aproximadamente de \$5. Para establecer las fronteras de cada clase, es necesario considerar la unidad más cercana con respecto a la cual se miden las observaciones. En este ejemplo las cuotas se presentan redondeadas al dólar más cercano. Con toda seguridad el importe de las cuotas es conocido hasta centavos, pero sólo se presentan los valores redondeados. Por ejemplo, la cuota de \$82 se interpreta como la media

entre $81.50 y $82.49, las seis clases con sus respectivas fronteras son (81.5-86.5), (86.5-91.5), (91.5-95.5), (96.5-101.5), (101.5-106.5) y (106.5-111.5).

Estas fronteras también se conocen como los *límites verdaderos* debido a que reflejan la unidad más pequeña que se emplea para tomar las observaciones. Dado que las cuotas se presentan redondeadas al dólar más cercano, se puede también elegir los límites de las seis clases como (82-86), (87-91), (92-96), (97-101), (102-106) y (107-111). Éstos se conocen como los *límites de escritura* puesto que reflejan el mismo grado de precisión que el de las observaciones presentadas. El intervalo de la clase es la diferencia entre los límites verdaderos de cada clase, mientras que los puntos medios pueden determinarse al utilizar los límites verdaderos o los de escritura. En la tabla 1.3 se da un resumen de la información pertinente para el agrupamiento de este ejemplo.

De acuerdo con lo mencionado al principio de esta sección, la distribución de frecuencia relativa se determina graficando las frecuencias relativas en el eje vertical contra los límites de escritura inferiores para cada una de las clases en el eje horizontal. Para este fin se emplean rectángulos de igual anchura que representen las frecuencias relativas. En la figura 1.3 se muestra el histograma del ejemplo 1.1. Nótese que es más fácil graficar las frecuencias de cada clase que las correspondientes frecuencias relativas; en ambos casos las gráficas serán idénticas. Si existe alguna preferencia para usar las frecuencias relativas, se debe a que la escala vertical tiene un intervalo fijo de cero a uno.

El principal objetivo de la representación gráfica de las frecuencias relativas es mostrar el perfil de distribución de los datos. El conocimiento de este perfil es útil en varias formas, como sugerían los análisis apropiados que se intentarán mediante la inferencia estadística, o si los datos constituyen una muestra aleatoria de alguna población o si se utilizan con el fin de comparar los perfiles de distribución de dos o más conjuntos de datos. En el ejemplo 1.1. es notorio que la distribución de cuotas anuales en las 40 compañías es uniforme a través de todo el intervalo de valores.

Otra caracterización gráfica útil, de un conjunto de datos, es la *distribución de frecuencia relativa acumulada* u *ojiva*. La distribución acumulativa se obtiene graficando, en el eje vertical, la frecuencia relativa acumulativa de una clase contra el

TABLA 1.3 Agrupamiento y frecuencias relativas para el ejemplo 1.1

Límites de escritura de la clase	*Punto medio*	*Frecuencia de la clase* f_i	*Frecuencia relativa* f_i/n
82–86	84	3	3/40 = 0.075
87–91	89	7	7/40 = 0.175
92–96	94	8	8/40 = 0.200
97–101	99	8	8/40 = 0.200
102–106	104	7	7/40 = 0.175
107–111	109	7	7/40 = 0.175
	Total	40*	1.000

$^*n = \sum_{i=1} f_i$

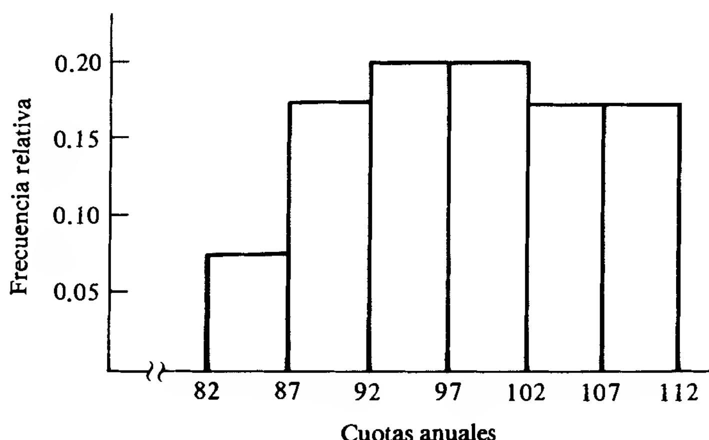


FIGURA 1.3 Distribución de frecuencia relativa para los datos del ejemplo 1.1

límite inferior de la siguiente sobre el eje horizontal y uniendo con segmentos todos los puntos consecutivos. La tabla 1.4 lista las frecuencias relativas acumuladas para el ejemplo 1.1.

Dado que la frecuencia relativa de una clase refleja la proporción de las observaciones contenidas en ésta, la frecuencia relativa acumulativa es la proporción de observaciones cuyos valores son menores o iguales al límite superior de la clase o, en forma equivalente, menores que el límite inferior de la siguiente clase. En el ejemplo 1.1 y para la tabla 1.4, la proporción de cuotas menores de $82 es cero. La de cuotas menores de $87 es de 0.075, la proporción de menores de $92 es de 0.250. La distribución de frecuencia relativa acumulativa para el ejemplo 1.1 se muestra en la figura 1.4.

En este contexto el principal uso de la distribución acumulativa es lo que comúnmente se conoce como cuantiles. Con respecto a una distribución de frecuencia relativa acumulativa, se define un *cuantil* como el valor bajo el cual se encuentra una determinada proporción de los valores de la distribución. El valor del cuantil se lee en

TABLA 1.4 Distribución de la frecuencia relativa acumulativa

Límites de escritura de la clase	*Frecuencia de clase*	*Frecuencia acumulativa*	*Frecuencia relativa acumulativa*
82–86	3	3	3/40 = 0.075
87–91	7	10	10/40 = 0.250
92–96	8	18	18/40 = 0.450
97–101	8	26	26/40 = 0.650
102–106	7	33	33/40 = 0.825
107–111	7	40	40/40 = 1.000

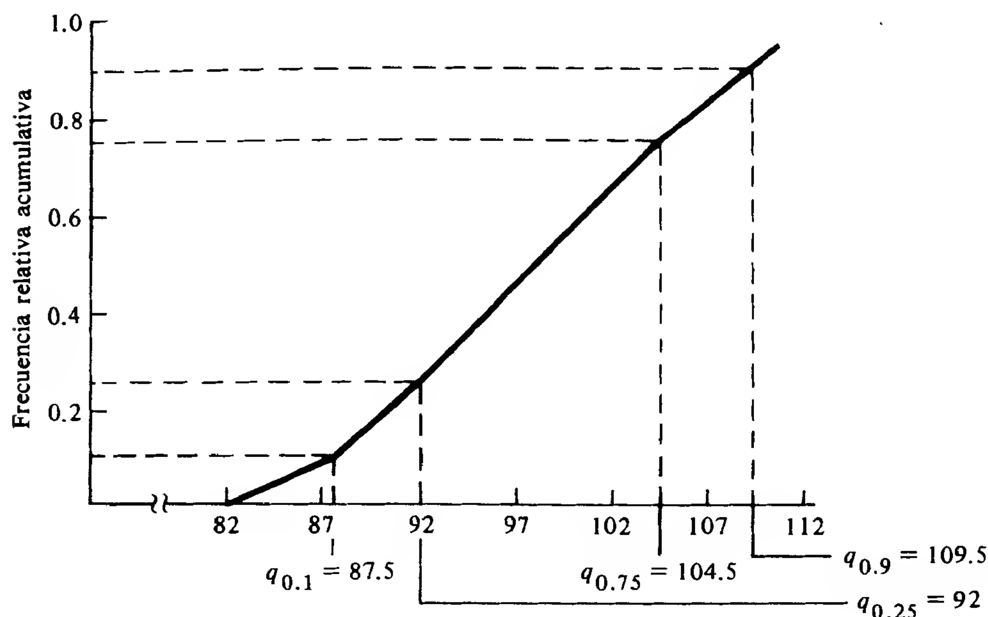


FIGURA 1.4 Distribución de frecuencia relativa acumulativa para el ejemplo 1.1

la dirección opuesta, en el eje horizontal, a la proporción correspondiente deseada sobre el eje vertical. El cuantil más común es el percentil. Por ejemplo, $q_{0.2}$ es el valor bajo el cual se encuentra el 20% de los valores de la distribución y $q_{0.9}$ es aquél bajo el cual se encuentra el 90% de los valores de la distribución.

Ejemplo 1.2 El departamento de Agricultura de Estados Unidos informó que, en 1976, los ingresos netos por cosecha para los 50 estados de la nación, fueron los siguientes:

$ 5 952	63 855	39 362	9 692	27 611
13 647	10 630	6 644	4 438	19 106
8 681	5 332	2 304	6 859	8 141
11 771	9 378	5 992	7 000	12 543
4 963	4 543	11 177	12 292	6 695
10 207	7 627	8 992	23 811	7 657
8 043	8 972	6 480	6 824	9 554
4 626	4 845	10 452	9 922	7 683
5 119	8 621	2 290	4 973	3 904
2 892	5 405	2 789	30	241

Establecer un esquema de agrupamiento para este conjunto de datos y determinar las frecuencias relativas.

TABLA 1.5 Frecuencias relativas para el ejemplo 1.2 con intervalos de igual longitud

Límites de escritura de la clase	*Frecuencia de la clase*	*Frecuencia relativa*
0–7 999	27	0.54
8 000–15 999	18	0.36
16 000–23 999	2	0.04
24 000–31 999	1	0.02
32 000–39 999	1	0.02
40 000–47 999	0	0
48 000–55 999	0	0
56 000–63 999	1	0.02
Total	50	1.00

Supóngase que se decide emplear ocho clases de igual longitud. Puesto que la diferencia entre los dos valores extremos del conjunto de datos es aproximadamente de $64 000, la longitud de cada clase es de $8 000 y los límites son (–0.5-7 999.5), (7 999.5-15 999.5), ..., (55 999.5-63 999.5). Las frecuencias de cada clase y las frecuencias relativas para este esquema de agrupamiento se dan en la tabla 1.5. Tal esquema resulta inadecuado porque el 90% de las observaciones se encuentran en las dos primeras clases y existen otras dos que no tienen ninguna observación. Este ejemplo ilustra un conjunto de datos para el que no deben usarse intervalos de igual longitud, ya que se tiene un agregado muy alto de observaciones con sólo algunas cuantas dispersas alrededor de éste. En el ejemplo 1.2 existe mayor concentración de datos en el extremo inferior que en el superior. Por consiguiente, considérese el siguiente esquema de agrupamiento de ocho clases con límites (–0.5-1 999.5), (1 999.5-3 999.5), (3 999.5-5 999.5), (5 999.5-7 999.5), (7 999.5-11 999.5), (11 999.5-27 999.5), (27 999.5-43 999.5), (43 999.5-75 999.5). La tabla 1.6 contiene las frecuencias relativas para este esquema, mientras que en la figura 1.5 se muestra la distribución de frecuencias.

Al determinar la distribución de frecuencia relativa de la figura 1.5, se empleó la altura del rectángulo en la representación de la frecuencia relativa de cada clase, de la misma manera como se hizo en el ejemplo 1.1. Sin embargo, a causa de que los intervalos no tienen la misma longitud, la figura 1.5 produce la impresión errónea de que, por ejemplo, la clase (12 000-27 999) contiene más del 12% de las observaciones. Lo anterior se debe a que cuando se comparan figuras geométricas, como los rectángulos, se tiende más a comparar el área que la altura. Cuando los intervalos de clase son idénticos, el área de los rectángulos representa las frecuencias. Sin embargo cuando la longitud de los intervalos es diferente, como en el ejemplo 1.2, las áreas no representan la frecuencia. Por lo tanto, es necesario ajustar la altura de los rectángulos para que sus áreas sean proporcionales a la frecuencia. Este procedimiento representa de manera correcta las frecuencias para intervalos de diferente longitud.

Para ilustrar este método, en el ejemplo 1.2, se observa que las longitudes de las primeras cuatro clases son idénticas. Entonces deben ajustarse las últimas cuatro con el fin de que sus longitudes se relacionen con las de las primeras cuatro clases (de $2 000). Las alturas de los rectángulos correspondientes a las cuatro últimas clases se

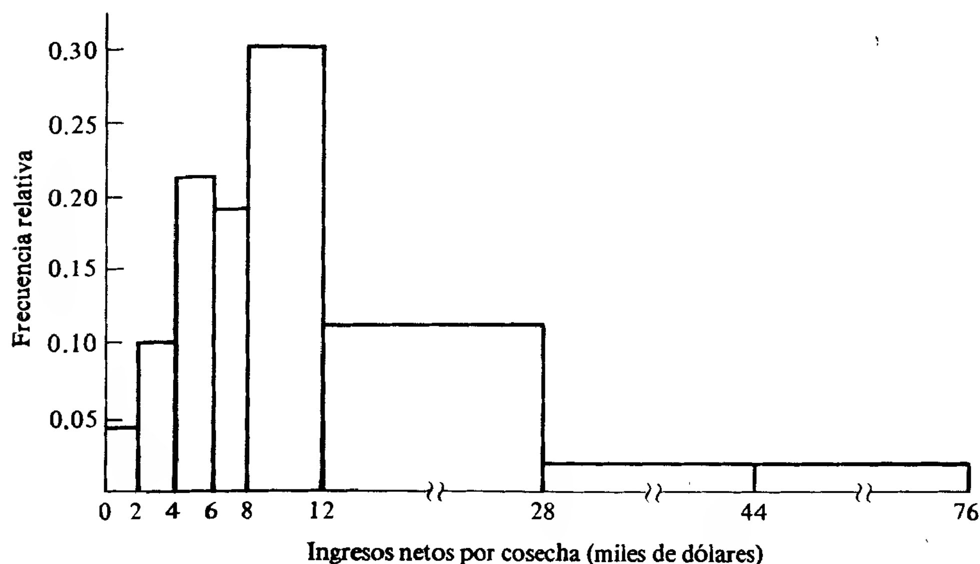


FIGURA 1.5 Distribución de frecuencia relativa para los ingresos por cosecha del año 1976

ajustan de tal forma que su área se encuentra en la misma proporción (2 000) con respecto a sus frecuencias relativas que las de los rectángulos de las primeras cuatro clases. Las alturas de las primeras cuatro siguen siendo las mismas que aparecen en la última columna de la tabla 1.6, mientras que las alturas corregidas para las últimas cuatro son 0.15, 0.015, 0.0025 y 0.00125 respectivamente. En este momento debe notarse que la suma de todas estas nuevas alturas es de 0.70875 y no de 1.00, como es requerido para frecuencias relativas. Una división por 0.70875 convertirá estas alturas a las frecuencias relativas deseadas. En la tabla 1.7 aparecen las frecuencias relativas corregidas y en la figura 1.6 se da la correcta representación de la distribución de frecuencia relativa.

TABLA 1.6 Frecuencias relativas para el ejemplo 1.2 con intervalos de distinta longitud

Límites de escritura de la clase	*Frecuencia de la clase*	*Frecuencia relativa*
0–1 999	2	0.04
2 000–3 999	5	0.10
4 000–5 999	11	0.22
6 000–7 999	9	0.18
8 000–11 999	15	0.30
12 000–27 999	6	0.12
28 000–43 999	1	0.02
44 000–75 999	1	0.02
Total	50	1.00

TABLA 1.7 Frecuencias relativas corregidas para el ejemplo 1.2 con intervalos de distinta longitud

Límites de escritura de la clase	*Frecuencia relativa corregida*
0–1 999	0.0564
2,000–3 999	0.1411
4,000–5 999	0.3104
6,000–7 999	0.2540
8,000–11 999	0.2116
12,000–27 999	0.0212
28,000–43 999	0.0035
44,000–75 999	0.0018
Total	1.0000

1.3 Medidas numéricas descriptivas

En la sección anterior se plantearon las técnicas gráficas para descubrir los patrones de distribución ocultos en un conjunto de datos. En esta sección se definen algunas medidas numéricas que se emplean comúnmente para describir conjuntos de datos. Si el conjunto es una muestra aleatoria de una población y la última meta es hacer inferencia estadística, estas medidas serán utilizadas como bases para las inferencias, tal como se menciona en los capítulos 7 a 9.

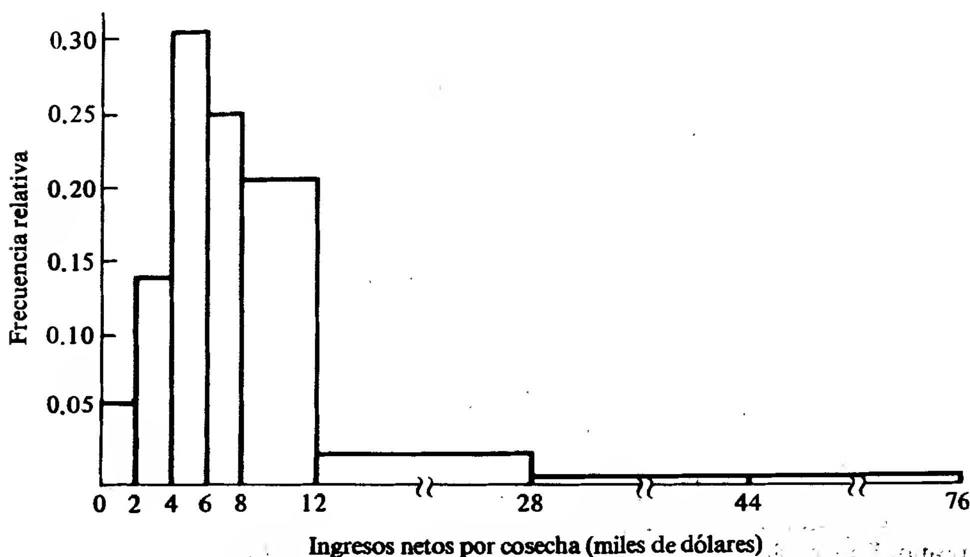


FIGURA 1.6 Distribución de frecuencia relativa corregida para los ingresos por cosecha del año 1976

Existen dos medidas de interés para cualquier conjunto de datos: la localización de su centro y su variabilidad. La *tendencia central* de un conjunto de datos es la disposición de éstos para agruparse ya sea alrededor del centro o de ciertos valores numéricos. La *variabilidad* de un conjunto de datos es la dispersión de las observaciones en el conjunto.

Existen principalmente tres medidas de tendencia central: la media, la mediana y la moda.

Definición 1.1 La *media* de las observaciones $x_1, x_2, \ldots, x_n$ es el promedio aritmético de éstas y se denota por

$$\bar{x} = \sum_{i=1}^{n} x_i/n. \tag{1.1}$$

La media es una medida apropiada de tendencia central para muchos conjuntos de datos. Sin embargo, dado que cualquier observación en el conjunto se emplea para su cálculo, el valor de la media puede afectarse de manera desproporcionada por la existencia de algunos valores extremos.

Definición 1.2. La *mediana* de un conjunto de observaciones es el valor para el cual, cuando todas las observaciones se ordenan de manera creciente, la mitad de éstas es menor que este valor y la otra mitad mayor.

Si el número de observaciones en el conjunto es impar, la mediana es el valor de la observación que se encuentra a la mitad del conjunto ordenado. Si el número es par se considera la mediana como el promedio aritmético de los valores de las dos observaciones que se encuentren a la mitad del conjunto ordenado. Alternativamente, la mediana puede determinarse a partir de la distribución acumulativa, es decir, la mediana es el percentil cincuenta.

Puesto que la mediana es un valor que se basa en la secuencia ordenada de las observaciones en un conjunto de datos, es necesario saber que la existencia de algunos valores extremos no afectará su valor. Por lo tanto, si un conjunto contiene unos cuantos valores extremos y un agregado muy alto de observaciones, la mediana puede ser una medida de tendencia central mucho más deseable que la media. Generalmente los conjuntos de datos que describen información acerca de ingresos caen en esta categoría.

Definición 1.3 La *moda* de un conjunto de observaciones es el valor de la observación que ocurre con mayor frecuencia en el conjunto.

La moda muestra hacia qué valor tienden los datos a agruparse. En conjuntos relativamente pequeños, puede que no exista un par de observaciones cuyo valor sea el mismo. En esta situación no es clara la definición de moda. También puede suceder que la frecuencia más alta se encuentre compartida por dos o más observaciones. En estos casos, la moda tiene una utilidad limitada como medida de tendencia central. Si se ha determinado una distribución de frecuencia relativa, la clase con la frecuen-

cia más alta recibirá el nombre de clase modal, con lo que se define a la moda como el punto medio de esa clase. En este caso la clase modal sirve como punto de concentración en el conjunto de datos.

Para las observaciones del ejemplo 1.1 la media se calcula como

$$\bar{x} = \frac{82 + 85 + \cdots + 111}{40} = \$97.90.$$

La media para el ejemplo 1.2 es

$$\bar{x} = \frac{5{,}952 + 63{,}855 + \cdots + 241}{50} = \$9\,811.34.$$

La mediana del ejemplo 1.1 es el promedio artimético de los valores de las observaciones 20 y 21 en la secuencia ordenada de éstas, ya que existe un número par de observaciones. La mediana es (98+99)/2 = $98.50. Similarmente, la mediana del ejemplo 1.2 es el promedio aritmético de los valores de las observaciones 25 y 26 en la secuencia ordenada de éstas, o (7 627 + 7 657)/2 = $7 642. Se observa que la moda en el ejempo 1.1 es $95 porque este valor es el que ocurre con mayor frecuencia; sin embargo, para el ejemplo 1.2 la moda no está claramente definida puesto que ningún valor se repite. Nótese que para el ejemplo 1.1 los valores de la media, mediana y moda se encuentran muy cercanos, relativamente, entre sí. Esto se debe a que las cuotas se encuentran distribuidas de manera uniforme sobre el intervalo completo de valores. Para el ejemplo 1.2 la media es sustancialmente mayor que la mediana, debido a que la primera se encuentra afectada de manera desproporcionada por los ingresos por cosecha de algunos estados, los que son muy grandes comparados con los de otros. Así, para este conjunto de datos la mediana de $7 642 podría ser una medida de tendencia central mucho más real.

Muchas veces la única información disponible es una tabla de frecuencias, como las tablas 1.3 a 1.6. En estos casos sólo es posible obtener valores aproximados para la media, mediana y moda — o para cualquier otra medida numérica descriptiva —; los valores exactos pueden calcularse únicamente a partir de las observaciones individuales del conjunto o de los datos no agrupados. Los cálculos aproximados se basan en los puntos medios de cada clase y sus respectivas frecuencias. En general, mientras más pequeña sea la longitud de la clase y mayor la uniformidad de las observaciones en ésta, mayor será la similitud entre las medidas descriptivas calculadas en los datos agrupados y no agrupados.

Para calcular la media con base en los datos agrupados, sea k el número de clases y x_i el punto medio de la i-ésima clase. Entonces el valor aproximado de la media es

$$\bar{x} = \sum_{i=1}^{k} f_i x_i / n, \tag{1.2}$$

en donde f_i es la frecuencia de la i-ésima clase y $n = \sum_{i=1}^{k} f_i$. Nótese que en esta fórmula la frecuencia de la clase representa la frecuencia relativa de las observaciones dentro de cada clase. Es decir, entre más observaciones tenga una clase mayor será el peso del punto medio de ésta en el cálculo de la media. La afirmación anterior gene-

TABLA 1.8 Cálculo aproximado de la media para el ejemplo 1.1

Punto medio de la clase x_i	*Frecuencia de la clase* f_i	$f_i x_i$
84	3	252
89	7	623
94	8	752
99	8	792
104	7	728
109	7	763
Total	40	3 910

$$n = \sum_{i=1}^{6} f_i = 40$$

$$\sum_{i=1}^{6} f_i x_i = 3\,910$$

$$\bar{x} = \sum_{i=1}^{6} f_i x_i / n = 3\,910/40 = \$97.75$$

ralmente es cierta en la determinación de medidas numéricas con base en datos agrupados.

Se ilustrarán los procedimientos computacionales para determinar las medidas descriptivas numéricas empleando el ejemplo 1.1 y en particular los límites y frecuencias de cada clase expuestos en la tabla 1.3. La información más importante aunada al cálculo de la media se muestra en la tabla 1.8.

Para datos agrupados, la mediana es aquel valor que divide en dos partes iguales la distribución de frecuencia relativa. La fórmula computacional está dada por

$$Mediana = L + c(j/f_m), \tag{1.3}$$

en donde L es el límite inferior de la clase donde se encuentra la mediana, f_m es la frecuencia de esa clase, c es la longitud de la clase y j es el número de observaciones en esta clase, necesarias para completar un total de $n/2$. Para determinar la mediana esta fórmula en esencia, se interpola linealmente en la clase que contiene a la mediana. Así, se supone que las observaciones se encuentran distribuidas uniformemente dentro de la clase.

La mediana para los datos agrupados del ejemplo 1.1 se determina utilizando la información contenida en la tabla 1.3. El número total de observaciones es 40 y $n/2$ es 20. Puesto que la suma de las frecuencias de las primeras tres clases es 18 y la de las primeras cuatro es 26, la mediana se encuentra en la cuarta clase, cuyo límite inferior es 97. Del total de observaciones en ésta clase, que es ocho, se necesitan dos más para alcanzar el valor de 20. Mediante el empleo de la fórmula, la mediana resulta ser

$$Mediana = 97 + 5(2/8) = \$98.25.$$

Como se mencionó anteriormente, la moda se toma, para datos agrupados, como el punto medio de la clase que presenta una mayor frecuencia. En el ejemplo 1.1 la frecuencia más alta se encuentra compartida por las clases (92-96) y (97-101). Con base en lo anterior, la moda resulta ser el promedio aritmético entre los dos puntos medios de las clases, o (94 + 92)/2 = \$96.50.

Una medida de tendencia central proporciona información acerca de un conjunto de datos pero no proporciona ninguna idea de la variabilidad de las observaciones

en dicho conjunto. Por ejemplo, considere los dos siguientes conjuntos de datos, cada uno de los cuales consiste de cuatro observaciones: 0, 25, 75, 100; 48, 49, 51, 52. En ambos casos, *media* = *mediana* = 50. Estos dos conjuntos son muy diferentes entre sí, sin embargo las observaciones en el primero se encuentran mucho más dispersas que las del segundo. Una de las medidas más útiles de dispersión o variación es la varianza.

Definición 1.4 La *varianza* de las observaciones $x_1, x_2, \cdots, x_n$ es, en esencia, el promedio del cuadrado de las distancias entre cada observación y la media del conjunto de observaciones. La varianza se denota por

$$s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/(n - 1). \tag{1.4}$$

La varianza es una medida razonablemente buena de la variabilidad debido a que si muchas de las diferencias son grandes (o pequeñas) entonces el valor de la varianza s^2 será grande (o pequeño). El valor de la varianza puede sufrir un cambio muy desproporcionado, aún más que la media, por la existencia de algunos valores extremos en el conjunto.

Definición 1.5 La raíz cuadrada positiva de la varianza recibe el nombre de *desviación estándar* y se denota por

$$s = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2/(n - 1)}. \tag{1.5}$$

La varianza y la desviación estándar no son medidas de variabilidad distintas, debido a que la última no puede determinarse a menos que se conozca la primera. A menudo se prefiere la desviación estándar en relación con la varianza, porque se expresa en las mismas unidades físicas de las observaciones.

Cuando se calcula el valor de la varianza, ya sea a mano o mediante el uso de una calculadora de baja capacidad, y el valor de la media o los valores de las observaciones no son números enteros, el uso de la ecuación (1.4) puede dar origen a errores grandes por redondeo. Con un poco de álgebra se obtiene, a partir de (1.4), una fórmula computacional más exacta para esas condiciones:*

$$s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/(n - 1)$$

$$= \frac{\sum(x_i^2 - 2\bar{x}\,x_i + \bar{x}^2)}{n - 1}$$

* **Para un repaso de la notación de suma véase el apéndice de este capítulo.**

$$= \frac{\sum x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2}{n - 1}$$

$$= \frac{\sum x_i^2 - \dfrac{2\left(\sum x_i\right)\left(\sum x_i\right)}{n} + \dfrac{n\left(\sum x_i\right)^2}{n^2}}{n - 1}$$

$$= \frac{\sum_{i=1}^{n} x_i^2 - \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n - 1}. \tag{1.6}$$

Nótese que para el numerador de la ecuación (1.4) primero debe calcularse la media, restarla de cada observación, tomar el cuadrado y entonces sumar. Para el numerador de (1.6) se suman todos los cuadrados de los valores observados, y entonces se resta el cuadrado de su suma dividido por el número de observaciones. Con base en la ecuación (1.6), la desviación estándar está dada por

$$s = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 \Big/ n}{n - 1}}. \tag{1.7}$$

A continuación se ilustran los pasos que se deben seguir para el cálculo de la varianza y la desviación estándar, para los datos no agrupados de los ejemplos 1.1 y 1 2. Para el ejemplo 1.1,

$$\sum_{i=1}^{40} x_i = 82 + 85 + \cdots + 111 = 3\,916$$

$$\sum_{i=1}^{40} x_i^2 = 82^2 + 85^2 + \cdots + 111^2 = 385\,756.$$

Se usa la ecuación (1.6),

$$s^2 = \frac{385\,756 - \dfrac{(3\,916)^2}{40}}{40 - 1} = 61.0154.$$

De la ecuación (1.7) se sigue que la desviación estándar es $s = \sqrt{61.0154} =$ \$7.81.

Para el ejemplo 1.2 se tiene

$$\sum_{i=1}^{50} x_i = 5\,952 + 63\,855 + \cdots + 241 = 490\,567,$$

$$\sum_{i=1}^{50} x_i^2 = 5\,952^2 + 63\,855^2 + \cdots + 241^2 = 10\,000\,514\,273,$$

y

$$s^2 = \frac{10\,000\,514\,273 - \dfrac{490\,567^2}{50}}{50 - 1} = 105\,865\,196.8.$$

La desviación estándar es $s = \$10\,289.08$.

Para datos agrupados, puede calcularse el valor aproximado de la varianza mediante el uso de la fórmula

$$s^2 = \frac{\sum_{i=1}^{k} f_i(x_i - \bar{x})^2}{n - 1} \qquad (1.8)$$

o

$$s^2 = \frac{\sum_{i=1}^{k} f_i x_i^2 - \dfrac{\left(\sum_{i=1}^{k} f_i x_i\right)^2}{n}}{n - 1} \qquad (1.9)$$

La fórmula para la desviación estándar es

$$s = \sqrt{\sum_{i=1}^{k} f_i(x_i - \bar{x})^2/(n - 1)}. \qquad (1.10)$$

Para las tres fórmulas anteriores f_i y x_i son, respectivamente, la frecuencia y el punto medio de la i-ésima clase, y n es la suma de todas las frecuencias. Debe notarse que, en datos agrupados, la aproximación a la varianza puede no ser muy confiable, especialmente si las observaciones no se encuentran distribuidas de manera uniforme dentro de sus respectivas clases. El cálculo de los valores aproximados de la varianza y la desviación estándar, para los datos agrupados del ejemplo 1.1, se encuentra detallado en la tabla 1.9.

Otra medida útil de la variabilidad tiene base en el valor absoluto de las diferencias entre las observaciones $x_1, x_2, \ldots x_n$ y la media o la mediana, dependiendo de cual de las dos se emplee como medida de tendencia central.

TABLA 1.9 Cálculo de los valores aproximados de la varianza y la desviación estándar para el ejemplo 1.1

Punto medio de la clase x_i	*Frecuencia de la clase* f_i	x_i^2	$f_i x_i^2$
84	3	7 056	21 168
89	7	7 921	55 447
94	8	8 836	70 688
99	8	9 801	78 408
104	7	10 816	75 712
109	7	11 881	83 167
Total	40	11 881	384 590

$$\sum_{i=1}^{6} f_i x_i = 3\,910 \text{ (de la tabla) 1.8)}$$

$$\left(\sum_{i=1}^{6} f_i x_i\right)^2 \Big/ 40 = 382\,202.5$$

$$\sum_{i=1}^{6} f_i x_i^2 = 384\,590$$

$$s^2 = \frac{384\,590 - 382\,202.5}{40 - 1}$$

$$= 61.2179$$

$$s = \sqrt{61.2179} = \$7.82$$

Definición 1.6 La *desviación media* es el promedio de los valores absolutos de las diferencias entre cada observación y la media de las observaciones. La desviación media está dada por

$$D.M. = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}. \tag{1.11}$$

Para datos agrupados, el valor de la desviación media se aproxima por

$$D.M. = \frac{\sum_{i=1}^{k} f_i |x_i - \bar{x}|}{\sum_{i=1}^{k} f_i}. \tag{1.12}$$

Los términos empleados en estas expresiones son los mismos definidos anteriormente.

La desviación media es una medida interesante de la variación, especialmente en el contexto de la evidencia empírica, debido a que en muchas ocasiones el interés se centra en las desviaciones y no en los signos de éstas. Sin embargo, desde un punto de vista teórico, el empleo de la desviación media como medida de dispersión está en desventaja dado que, matemáticamente, es difícil de obtener. De cualquier manera, la desviación media es menos sensible a los efectos inducidos por las observaciones extremas del conjunto de datos que la varianza o la desviación estándar. Sin importar la presencia de pocos valores extremos, la desviación media puede proporcionar una medida de dispersión mucho más real que la obtenida por la desviación estándar.

Para los datos no agrupados del ejemplo 1.1, la desviación media se calcula a partir de

$$\sum_{i=1}^{40} |x_i - \bar{x}| = |82 - 97.9| + |85 - 97.9| + \cdots + |111 - 97.9| = 264.2$$

para ser

$$D.M. = 264.2/40 = \$6.61.$$

De manera similar para el ejemplo 1.2, la desviación media se calcula a partir de

$$\sum_{i=1}^{50} |x_i - \bar{x}| = |5\,952 - 9\,811.34| + |63\,855 - 9\,811.34| + \cdots + |241 - 9\,811.34|$$

$$= 278\,051.48$$

para ser

$$D.M. = 278\,051.48/50 = \$5\,561.03.$$

Los pasos computacionales para una aproximación de la desviación media a los datos agrupados del ejemplo 1.1, se ilustran en la tabla 1.10.

Definición 1.7 La *desviación mediana* es el promedio de los valores absolutos de las diferencias entre cada observación y la mediana de éstas. La desviación mediana está dada por

$$D.Md. = \frac{\sum_{i=1}^{n} |x_i - D.Md|}{n}, \tag{1.13}$$

en donde *Md* denota a la mediana.

Cuando la mediana se emplea como medida de tendencia central con el propósito de atenuar los efectos de la existencia de algunos valores extremos en el conjunto,

TABLA 1.10 Cálculo aproximado de la desviación de la mediana para el ejemplo 1.1

Punto medio de la clase x_i	*Frecuencia de la clase* f_i	$\lvert x_i - \bar{x}\rvert$	$f_i\lvert x_i - \bar{x}\rvert$
84	3	$\lvert 84 - 97.75\rvert$	41.25
89	7	$\lvert 89 - 97.75\rvert$	61.25
94	8	$\lvert 94 - 97.75\rvert$	30.00
99	8	$\lvert 99 - 97.75\rvert$	10.00
104	7	$\lvert 104 - 97.75\rvert$	43.75
109	7	$\lvert 109 - 97.75\rvert$	78.75
Total	40		265.00

$$\sum_{i=1}^{6} f_i|x_i - \bar{x}| = 265$$

$$D.M. = 265/40$$

$$= \$6.63$$

debe preferirse a la desviación de la mediana como medida de dispersión por la misma razón. Cuando los datos se agrupan, se obtiene el valor aproximado de la desviación de la mediana al emplear la ecuación (1.12) y sustituir la mediana por la media. Las desviaciones de las medianas para las observaciones de los ejemplos 1.1 y 1.2 calculadas con el mismo procedimiento que para las desviaciones de las medias, son 6.6 y 5 060.60 respectivamente. De manera similar el valor aproximado de la desviación de la mediana para los datos agrupados del ejemplo 1.1 tiene un valor de 6.575.

El intervalo en el que se encuentran las observaciones en un conjunto de datos, es otra medida de variabilidad.

Definición 1.8 El *recorrido R* de las observaciones en un conjunto de datos es la diferencia entre el valor más grande y el más pequeño del conjunto.

Por su simplicidad, el recorrido proporciona una rápida indicación de la variabilidad existente entre las observaciones de un conjunto de datos. Sin embargo, como medida de dispersión debe usarse con precaución ya que su valor es una función, únicamente, de dos valores extremos pertenecientes al conjunto. Como regla general se debe evitar el uso del recorrido como medida de variabilidad, cuando el número de observaciones en un conjunto es grande o cuando éste contenga algunas observaciones cuyo valor sea relativamente grande. Este punto puede ilustrarse considerando los recorridos de los ejemplos 1.1 y 1.2, que son $R_1 = 111 - 82 = \$29$, y $R_2 = 63\,855 - 30 = \$63\,825$, respectivamente. Para el ejemplo 1.1, R_1 parece ser una medida realista de la variabilidad, debido principalmente a que el conjunto no contiene ninguna cuota que se salga de la línea relativa a las otras. Sin embargo, para el ejemplo 1.2, R_2 no es una medida realista de la variabilidad, dado que los valores de \$30 y \$63 855 son, aparentemente, valores extremos con respecto a los ingresos netos por cosecha de gran parte de los otros estados. Para muchos problemas tiene una mayor utilidad determinar el recorrido entre dos valores cuantiles que entre dos valores extremos.

Definición 1.9 La diferencia entre los percentiles 75avo y 25avo recibe el nombre de *recorrido intercuantil*.

Definición 1.10 La diferencia entre los percentiles 90avo y décimo recibe el nombre de *recorrido interdecil*.

El recorrido intercuantil refleja la variabilidad de las observaciones comprendidas entre los percentiles 25 y 75 en el conjunto de datos, y el recorrido interdecil indica la dispersión de las observaciones con valores entre los percentiles 90 y 10. El resultado es que ni el rango intercuantil ni el interdecil son afectados por la presencia de observaciones relativamente grandes.

Para datos agrupados se pueden aproximar los recorridos intercuantil e interdecil a partir de la distribución de frecuencia relativa acumulada. Para ilustrar, empleando la figura 1.1, los valores aproximados de los rangos intercuantil e interdecil para el ejemplo 1.1 son $q_{0.75} - q_{0.25} = 104.50 - 92 = \$12.50$, y $q_{0.9} - q_{0.1} = 109.5 - 87.5 = \$22$, respectivamente. Para un conjunto de datos no agrupados

que contenga n observaciones, los percentiles 75avo y 25avo son los valores de las observaciones cuyos números de posición en la secuencia ordenada de observaciones, corresponden a $0.75n + 0.5$ y $0.25n + 0.5$, respectivamente. De manera similar, los percentiles 90 y décimo corresponden a los valores de las observaciones cuyos números de posición, con respecto a la secuencia ordenada, son $0.9n + 0.5$ y $0.1n + 0.5$ respectivamente. Para los datos del ejemplo 1.2, los percentiles 25 y 75 son los valores de las observaciones 13 y 38 correspondientes a la secuencia ordenada de las observaciones, respectivamente. De esta manera, $q_{0.25}$ = \$4 973, $q_{0.75}$ = \$10 207, siendo el recorrido intercuantil de \$5 234. Dado que para $n = 50$ $0.1n + 0.5 = 5.5$, el décimo percentil es el promedio de los valores 5 y 6, de las observaciones ordenadas, o $q_{0.1}$ = 2 840.5. Similarmente el percentil 90avo es el promedio de las observaciones 45 y 46 correspondientes a la secuencia ordenada, o $q_{0.9}$ = 16 376.5. Por lo tanto, el recorrido interdecil para los datos del ejemplo 1.1 es de \$13 536.

A lo largo de todo el capítulo se han empleado los ejemplos 1.1 y 1.2 para ilustrar varios conceptos. Es importante notar que presentan situaciones contrastantes. El primero presenta un conjunto de datos en el que las observaciones se encuentran distribuidas de manera uniforme a lo largo del recorrido completo de valores, sin ninguna observación relativamente grande. El último ejemplifica una situación en la que existe un agregado muy denso de observaciones y algunos valores relativamente grandes, especialmente en el extremo superior. La diferencia innata entre estos dos ejemplos, puede discernirse a través de una comparación de las medidas descriptivas numéricas que se han calculado para cada uno de ellos y que aparecen en la tabla 1.11.

Nótese que en el ejemplo 1.1 los valores de las medidas de tendencia central se encuentran muy cercanos entre sí, mientras que para el ejemplo 1.2 se encuentran separadas entre sí de manera considerable. Se puede decir lo mismo de las desviaciones estándar, media y mediana para los dos ejemplos. En el ejemplo primero los valores de las desviaciones de la media y de la mediana se encuentran muy próximos al valor de la desviación estándar, mientras que en el ejemplo 1.2 tienen un valor casi similar a la mitad de la desviación estándar. Además, en el ejemplo 1.1 el recorrido interdecil constituye una proporción relativamente grande del recorrido (22/29 = 0.76),

TABLA 1.11 Resumen de las medidas numéricas descriptivas para los ejemplos 1.1 y 1.2

Medida	*Ejemplo 1.1*		*Ejemplo 1.2*
numérica	*Datos no agrupados*	*Datos agrupados*	*Datos no agrupados*
Media	97.90	97.75	9 811.34
Mediana	98.50	98.25	7 642.00
Moda	95.00	96.50	—
Varianza	61.0154	61.2179	105 865 196.80
Desviación estándar	7.81	7.82	10 289.08
Desviación media	6.61	6.63	5 561.03
Desviación mediana	6.60	6.575	5 060.60
Recorrido	29.00	—	63 825.00
Recorrido intercuantil	—	12.50	5 234.00
Recorrido interdecil	—	22.00	13 536.00

y en el ejemplo 1.2 esta medida es una porción relativamente pequeña de este último (13 536/63 825 = 0.21).

Estas comparaciones aclaran lo que las medidas numéricas y las distribuciones de frecuencia pueden hacer para descubrir la naturaleza inherente de un conjunto de datos. Sin embargo, el usuario debe tener cuidado tanto en la elección como en la interpretación de estas medidas. A pesar de que la media y la desviación estándar se han empleado de manera extensa como medidas de tendencia central y dispersión respectivamente, aunque tienen propiedades teóricas muy atractivas existen problemas — como el ejemplo 1.2 — para los cuales no pueden ser las medidas más deseables. En general, y por ende, las medidas más deseables para conjuntos de datos relacionados con mediciones físicas como lecturas de instrumentos, especificaciones de partes, pesos, etc., son la medida y la desviación estándar o la desviación de la mediana. Para conjuntos de datos relacionados con ingresos y otras informaciones de tipo económico y financiero, las mejores elecciones para las medidas de tendencia central y dispersión son la mediana y la desviación de la mediana respectivamente.

Como nota final, las agencias del gobierno y muchos servicios de información proporcionan información en tablas de frecuencia que no sólo contienen clases de amplitud diferente sino también clases abiertas como "ingreso anual de $500 000 o más" con el propósito de tener mayor cobertura de los datos. Estas clases se presentan en los extremos del conjunto y no se especifican las clases terminales. Como resultado, el punto medio de las clases abiertas no se encuentra definido y no pueden calcularse valores aproximados para algunas medidas numéricas como la media, varianza, desviación estándar y desviación media, a menos que se encuentren disponibles algunas observaciones individuales contenidas en la clase o que sea conocido su promedio artimético.

Referencia

1. N.L. Johnson y F.C. Leone, *Statistics and experimental design,* Vol. I, segunda edición, Wiley, New York, 1977.

Ejercicios

1.1. Los siguientes datos son los lapsos, en minutos, necesarios para que 50 clientes de un banco comercial, lleven a cabo una transacción bancaria:

2.3	0.2	2.9	0.4	2.8
2.4	4.4	5.8	2.8	3.3
3.3	9.7	2.5	5.6	9.5
1.8	4.7	0.7	6.2	1.2
7.8	0.8	0.9	0.4	1.3
3.1	3.7	7.2	1.6	1.9
2.4	4.6	3.8	1.5	2.7
0.4	1.3	1.1	5.5	3.4
4.2	1.2	0.5	6.8	5.2
6.3	7.6	1.4	0.5	1.4

a) Construir una distribución de frecuencia relativa.
b) Construir una distribución de frecuencia relativa acumulada.
c) Con los resultados de la parte *b,* determine los recorridos intercuantil e interdecil.
d) Con los datos agrupados, calcule la media, mediana, moda, desviación estándar, desviación media y desviación mediana.
e) Verificar los resultados de la parte *d* calculando las mismas medidas para los datos no agrupados.

1.2. La demanda diaria, en unidades de un producto, durante 30 días de trabajo es:

38	35	76	58	48	59
67	63	33	69	53	51
28	25	36	32	61	57
49	78	48	42	72	52
47	66	58	44	44	56

a) Construir las distribuciones de frecuencia relativa y de frecuencia acumulada.
b) Con la distribución acumulada, determine los tres cuantiles.
c) Calcular la media, mediana, moda, desviación estándar, desviación media y desviación mediana, empleando tanto los datos agrupados como los no agrupados, y compare los dos conjuntos de resultados.
d) Comentar la naturaleza de esta distribución de frecuencia, cuando se compara con la del ejercicio 1.1.

1.3. Aquí se presentan tres conjuntos de datos:

$$1, 2, 3, 4, 5, 6;$$
$$1, 1, 1, 6, 6, 6;$$
$$-13, 2, 3, 4, 5, 20.$$

Calcular la media y la varianza para cada conjunto de datos. ¿Qué se puede concluir?

1.4. La siguiente tabla muestra las ventas, en miles de dólares, de 20 vendedores de una compañía de computadoras.

40.2	29.3	35.6	88.2	42.9
26.9	28.7	99.8	35.6	37.8
44.2	32.3	55.2	50.6	25.4
31.7	36.8	45.2	25.1	39.7

a) Calcular la media, mediana, desviación estándar, desviación mediana, recorrido intercuantil y recorrido interdecil.
b) ¿Qué medidas de tendencia central y dispersión se elegirían y por qué?

1.5. Con los datos del ejercicio 1.2, sea x_i la demanda del i-ésimo día para $i = 1, 2 \ldots 30$. Transformar los datos por medio de la relación

$$u_i = \frac{x_i - 51.5}{14.17}$$

a) Construir una distribución de frecuencia relativa para los datos transformados. ¿Ha ocurrido algún cambio en la naturaleza de la distribución de frecuencia cuando ésta se compara con la del ejercicio 1.2?

b) Con los datos transformados u_i, calcular la media y la desviación estándar; mostrar que son iguales a cero y uno respectivamente.

1.6. Los siguientes datos agrupados representan los pagos por almacenamiento para los 50 más grandes detallistas durante el año 1979:

Límites de estructura de la clase	*Frecuencia*
1.10–1.86	4
1.87–2.63	14
2.64–3.40	11
3.41–4.17	9
4.18–4.94	7
4.95–5.71	1
5.72–6.48	2
6.49–7.25	2

a) Graficar la distribución de frecuencia relativa acumulada.

b) Con los resultados de la parte *a)*, determinar los recorridos intercuantil e interdecil.

c) Calcular la media, mediana y moda.

d) Calcular la varianza, desviación estándar, desviación media y desviación mediana.

1.7. La siguiente información agrupada representa el número de puntos anotados por equipo y por juego en la Liga Nacional de Fútbol durante la temporada de 1973:

Grupo	*Frecuencia*
0–3	27
4–10	66
11–17	91
18–24	70
25–31	57
32–38	34
39–45	16
46–52	3

a) Graficar la distribución de frecuencia relativa.

b) Calcular la media y la moda.

c) Calcular la varianza, desviación estándar y desviación media.

1.8. Se seleccionaron de un proceso de fabricación, aleatoriamente, 20 baterías y se llevó a cabo una prueba para determinar la duración de éstas. Los siguientes datos representan el tiempo de duración, en horas, para las 20 baterías:

52.5	62.7	58.9	65.7	49.3
58.9	57.3	60.4	59.6	58.1
62.3	64.4	52.7	54.9	48.8
56.8	53.1	58.7	61.6	63.3

a) Determinar la media y la mediana.

b) Determinar la desviación estándar, desviación media y desviación mediana.

c) Determinar los recorridos intercuantil e interdecil.

APÉNDICE

Sumatorias y otras notaciones simbólicas

El uso de la notación simbólica es esencial en estadística. Por ejemplo, para distinguir entre los valores de n observaciones se emplea la notación simbólica $x_1, x_2, \ldots, x_n$. Uno de los símbolos más útiles es la letra griega Σ (sigma) con que se denota la suma de términos en una secuencia. De esta manera la suma de $x_1, x_2 \ldots, x_n$ se designa por

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n,$$

y se lee "la suma de las x_i, con i variando desde 1 hasta n". La letra i recibe el nombre de *índice de suma* y toma valores enteros sucesivos hasta e incluyendo a n, que es el límite superior o el valor más grande de i. Los siguientes son ejemplos del uso de Σ

a) $$\sum_{i=1}^{n} x_i^2 = x_1^2 + x_2^2 + \cdots + x_n^2;$$

b) $$\sum_{i=1}^{n} (x_i - a) = (x_1 - a) + (x_2 - a) + \cdots + (x_n - a);$$

c) $$\sum_{i=1}^{n} (x_i - a)^2 = (x_1 - a)^2 + (x_2 - a)^2 + \cdots + (x_n - a)^2;$$

d) $$\sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

Las siguientes tres propiedades son importantes cuando se emplea el símbolo Σ,

Propiedad 1. Si c es cualquier constante, entonces

$$\sum_{i=1}^{n} c x_i = c \sum_{i=1}^{n} x_i.$$

Propiedad 2. Si c es cualquier constante, entonces

$$\sum_{i=1}^{n} c = nc.$$

Propiedad 3.

$$\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i.$$

Las propiedades anteriores pueden verificarse de la siguiente manera:

$$1)\ \sum_{i=1}^{n} cx_i = cx_1 + cx_2 + \cdots + cx_n$$

$$= c(x_1 + x_2 + \cdots + x_n)$$

$$= c \sum_{i=1}^{n} x_i.$$

$$2)\ \sum_{i=1}^{n} c = \underbrace{c + c + \cdots + c}_{n \text{ términos}}$$

$$= \underbrace{(1 + 1 + \cdots + 1)c}_{n \text{ términos}}$$

$$= nc.$$

$$3)\ \sum_{i=1}^{n}(x_i + y_i) = (x_1 + y_1) + (x_2 + y_2) + \cdots + (x_n + y_n)$$

$$= (x_1 + x_2 + \cdots + x_n) + (y_1 + y_2 + \cdots + y_n)$$

$$= \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i.$$

El símbolo Σ también se emplea para denotar la suma sobre dos características diferentes. Por ejemplo, supóngase que se tiene la función $p(x, y)$ de las variables x y y, las que toman únicamente valores enteros. En particular x toma los valores enteros de 0 y 1, y y valores 1, 2 y 3. Entonces la suma de $p(x, y)$ sobre todos los valores tanto de x como de y se denota por

$$\sum_{x=0}^{1} \sum_{y=1}^{3} p(x, y) = p(0, 1) + p(0, 2) + p(0, 3) + p(1, 1) + p(1, 2) + p(1, 3).$$

Nótese que primero se elige el índice de suma de x igual a cero y entonces se evalúa la suma interna para cada uno de los valores del índice de suma de y. Posteriormente se incrementa el índice de suma de x en uno y se repite el proceso. El procedimiento anterior también se aplica a todas aquellas situaciones en las que se emplean subscritos dobles para distinguir entre dos características. Por ejemplo, considere la suma de la secuencia x_{ij}, $i = 1, 2 \ldots n$, $j = 1, 2 \ldots m$ para todos los valores posibles de i y de j. Tal suma puede denotarse por

$$\sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij}.$$

En particular, si $n = 2$ y $m = 3$, entonces

$$\sum_{i=1}^{2} \sum_{j=1}^{3} x_{ij} = x_{11} + x_{12} + x_{13} + x_{21} + x_{22} + x_{23}.$$

Otro símbolo últil es la letra griega Π (pi). Esta letra se emplea para indicar el producto de los términos de una secuencia. Por ejemplo, dada la secuencia de observaciones $x_1, x_2, \ldots, x_n$, el producto de $x_1, x_2, \ldots, x_n$ se denota por

$$\prod_{i=1}^{n} x_i = x_1 x_2 \ldots x_n,$$

en donde la letra i tiene el mismo propósito que en la suma.

CAPÍTULO DOS

Conceptos en probabilidad

2.1 Introducción

La probabilidad es un mecanismo por medio del cual pueden estudiarse sucesos aleatorios, cuando éstos se comparan con los fenómenos determinísticos. Por ejemplo, nadie espera predecir con certidumbre el resultado de un experimento tan simple como el lanzamiento de una moneda. Sin embargo, cualquier estudiante de primer año de licenciatura en física debe ser capaz de calcular el tiempo que transcurrirá para que un objeto, que se deja caer desde una altura conocida, llegue al suelo.

La probabilidad tiene un papel crucial en la aplicación de la inferencia estadística porque una decisión, cuyo fundamento se encuentra en la información contenida en una muestra aleatoria, puede estar equivocada. Sin una adecuada comprensión de las leyes básicas de la probabilidad, es difícil utilizar la metodología estadística de manera efectiva.

Para ilustrar el uso de la probabilidad en la toma de decisiones, considérese el siguiente ejemplo: una compañía produce un detergente líquido que se envasa en botellas de 500 ml, las que son llenadas por una máquina. Debido a que las botellas que contienen una cantidad mayor de 500 ml representan una pérdida para la compañía y todas aquellas que contienen una cantidad menor constituyen una pérdida para el consumidor (lo que puede desencadenar una acción legal en contra de la compañía), la compañía realiza todos los esfuerzos necesarios para mantener el volumen neto promedio en un nivel de 500 ml. Para mantener un control apropiado se ideó el siguiente esquema de muestreo: se seleccionarán 10 botellas del proceso de llenado, cuatro veces durante el transcurso del día y se determinará su contenido neto promedio. Si éste se encuentra entre 498 y 502 ml, inclusive, el proceso se considerará "bajo control"; de otra manera, éste se encontrará "fuera de control". En este caso se detendrá el llenado, llevando a cabo todos los esfuerzos necesarios para determinar la causa, si es que ésta existe, del problema. Con toda seguridad y para cualquiera de las dos situaciones se tienen riesgos. Si el proceso se considera bajo control, podría encontrarse fuera de éste, y la compañía puede estar perdiendo el producto o sujetándose a una acción legal por parte de las correspondientes oficinas del gobierno. Por otro lado si el proceso se considera fuera de control, puede en realidad encontrarse bajo control y la compañía estará intentando localizar una falla

inexistente. La evaluación de estos riesgos sólo puede hacerse de manera efectiva a través del uso de la probabilidad.

En las tres secciones siguientes se examinarán las interpretaciones *clásica, de frecuencia relativa* y *subjetiva,* de la probabilidad. Las dos primeras son muy similares debido a que se basan en la repetición de experimentos realizados bajo las mismas condiciones, como el lanzamiento de una moneda. La interpretación subjetiva o personal de la probabilidad representa una medida del grado de creencia con respecto a una proposición, como podría ser si la creación de una nueva empresa tendrá éxito. En la sección 2.5 se establecen algunos axiomas y, con base en éstos, se define formalmente la probabilidad. El desarrollo axiomático incluye las tres interpretaciones de la probabilidad.

2.2 La definición clásica de probabilidad

El desarrollo inicial de la probabilidad se asocia con los juegos de azar. Por ejemplo, considérense dos dados que se distingan y que no están cargados; el interés recae en los números que aparecen cuando se tiran los dados. En la tabla 2.2 se dan los 36 posibles pares de números.

Una característica clave de este ejemplo, así como también de muchos otros relacionados con los juegos de azar, es que los 36 resultados son *mutuamente excluyentes* debido a que no puede aparecer más de un par en forma simultánea. Los 36 resultados son *igualmente probables* puesto que sus frecuencias son prácticamente las mismas, si se supone que los dados no están cargados y que el experimento se lleva a cabo un número suficientemente grande de veces. Nótese que de los 36 resultados posibles, seis dan una suma de siete, cinco dan una suma de ocho, etc. Por lo tanto, puede pensarse de manera intuitiva que la probabilidad de obtener un par de números cuya suma sea siete es la proporción de resultados que suman siete con respecto al número total, en este caso 6/36. Es importante que el lector comprenda que la proporción 6/36 se obtiene únicamente después de que el experimento se realiza un número grande de veces, es decir, después de efectuar el experimento muchas veces se observará que, alrededor de la sexta parte de éste, la suma de los números que aparecen es igual a siete. La proporción 6/36 no significa que en seis tiradas, forzosamente una dará como resultado un siete. Para situaciones de este tipo es apropiada la siguiente definición de probabilidad.

Definición 2.1 Si un experimento que está sujeto al azar, resulta de n formas igualmente probables y mutuamente excluyentes, y si n_A de estos resultados tienen un atributo A, la probabilidad de A es la proporción de n_A con respecto a n.

TABLA 2.1 Posibles resultados que aparecen cuando se lanzan dos dados

1,1	1,2	1,3	1,4	1,5	1,6
2,1	2,2	2,3	2,4	2,5	2,6
3,1	3,2	3,3	3,4	3,5	3,6
4,1	4,2	4,3	4,4	4,5	4,6
5,1	5,2	5,3	5,4	5,5	5,6
6,1	6,2	6,3	6,4	6,5	6,6

2.3 Definición de probabilidad como frecuencia relativa

En muchas situaciones prácticas, los posibles resultados de un experimento no son igualmente probables. Por ejemplo, en una fábrica las oportunidades de observar un artículo defectuoso normalmente será mucho más rara que observar un artículo bueno. En este caso, no es correcto estimar la probabilidad de encontrar un artículo defectuoso mediante el empleo de la definición clásica. En lugar de ésta, en muchas ocasiones se emplea la interpretación de la probabilidad como una frecuencia relativa.

La interpretación de una frecuencia relativa descansa en la idea de que un experimento se efectúa y se repite muchas veces, y prácticamente bajo las mismas condiciones. Cada vez que un experimento se lleva a cabo, se observa un resultado. Éste es impredecible dada la naturaleza aleatoria del experimento, la probabilidad de la presencia de cierto atributo se aproxima por la frecuencia relativa de los resultados que posee dicho atributo. Conforme aumenta la repetición del experimento, la frecuencia relativa de los resultados favorables se aproxima al verdadero valor de la probabilidad para ese atributo. Por ejemplo: supóngase que se desea determinar la proporción de artículos defectuosos en un proceso de fabricación. Para llevar a cabo lo anterior, se muestra un determinado número de artículos; cada observación constituye un experimento. Los resultados pueden clasificarse como defectuosos o no defectuosos. Si el proceso de fabricación es estable, y asegura así las condiciones uniformes, al aumentar el número de artículos muestreados, la frecuencia relativa de artículos defectuosos con respecto al número de unidades muestreadas se aproximará cada vez más a la verdadera proporción de artículos defectuosos.

Para ilustrar la interpretación de la probabilidad como frecuencia relativa se simuló en una computadora un proceso de muestreo de n unidades, suponiendo que el proceso de fabricación producía un 5% de artículos defectuosos. Para cada n se observó el número de unidades defectuosas; los resultados se dan en la tabla 2.2 para valores de n entre 20 y 10 000. A partir de esto es razonable concluir que la frecuencia relativa tiende a un valor verdadero de 0.05 conforme n crece. De esta manera, se sugiere la siguiente definición de la probabilidad como frecuencia relativa:

TABLA 2.2. Resultados de un experimento simulado en computadora

Número de unidades muestreadas (n)	*Número de unidades defectuosas observadas*	*Frecuencia relativa*
20	2	0.10
50	3	0.06
100	4	0.04
200	12	0.06
500	28	0.056
1 000	54	0.054
2 000	97	0.0485
5 000	244	0.0488
10 000	504	0.0504

Definición 2.2 Si un experimento se repite n veces bajo las mismas condiciones y n_B de los resultados son favorables a un atributo B, el límite de n_B/n conforme n se vuelve grande, se define como la probabilidad del atributo B.

2.4 Interpretación subjetiva de la probabilidad

La repetición de un experimento bajo las mismas condiciones es la base para las interpretaciones clásica y de frecuencia relativa de la probabilidad. Sin embargo, muchos fenómenos no se prestan para repetición, pero a pesar de esto requieren de una noción de probabilidad. Por ejemplo la compañía que aseguró los Juegos Olímpicos de 1980 tuvo que determinar, *a priori*, los riesgos de que los juegos no se efectuasen de la manera en que se habían planeado. O cuando se aseguran contra robo o daño esculturas y pinturas cuyo valor es muy alto, las compañías aseguradoras deben tener idea de los riesgos adquiridos para fijar de manera adecuada, el precio del seguro. En ninguno de estos ejemplos puede concebirse un experimento susceptible de llevarse a cabo bajo condiciones similares. Por otra parte, muchas de las afirmaciones que suelen formularse las personas de algún modo implican probabilidad. Por ejemplo, cuando se dice "probablemente el embarque llegará mañana", o cuando un corredor de bolsa asesora a un cliente sobre la posible alza de una acción, se está sugiriendo alguna idea de la probabilidad de ocurrencia de las afirmaciones anteriores.

Para los ejemplos anteriores, la interpretación de la probabilidad no puede tener su fundamento en la frecuencia de ocurrencia. La probabilidad se interpreta como el grado de creencia o de convicción con respecto a la ocurrencia de una afirmación. En este contexto, la probabilidad representa un juicio personal acerca de un fenómeno impredecible. Esta interpretación de la probabilidad se conoce como *subjetiva* o *personal*.

Es importante hacer hincapié en que la probabilidad subjetiva también puede aplicarse a experimentos repetitivos. Por ejemplo, un jugador de blackjack puede, en un momento dado, decidir tomar otra carta y hacer caso omiso de su experiencia previa, debido a que cree que esto aumentará sus oportunidades de ganar la mano. El capitán de un equipo de futbol puede pedir "cara" cuando la moneda se lance al aire, debido a que ésa es su creencia con respecto al resultado de arrojarla. Con base en tales aplicaciones, la probabilidad subjetiva es considerada por muchos como más general que las otras dos interpretaciones.

Para ilustrar la traslación de un grado de creencia en probabilidad, considere la siguiente situación: se pregunta a dos ingenieros petroleros, A y B, su opinión acerca de la posibilidad de descubrir petróleo en un determinado sitio. La respuesta de A es que él está seguro, en un 80%, de que se encontrará petróleo mientras que B lo está en un 70%.* El porcentaje dado por los ingenieros es una medida de la creencia de éstos, con respecto al descubrimiento de petróleo. De esta manera se pueden asignar distintas medidas de creencia a la misma proposición. Pero ¿qué significado tienen realmente el 80% y 7%? La interpretación común es la siguiente. El ingeniero A pien-

* Por implicación, A y B también están diciendo que se encuentran seguros, en un 20% y 30%, respectivamente, de que no será descubierto el petróleo.

sa apostar ocho a dos (por ejemplo $8 contra $2 o cualquier otra cantidad de dólares que se encuentre en la misma proporción) a que el petróleo será descubierto en ese sitio. De manera similar, B cree que es mejor apostar siete a tres (es decir $7 contra $3) para el mismo resultado. De esta manera, las probabilidades subjetivas de A y B se definen como las proporciones $8/(8 + 2)$ y $7/(7 + 3)$ respectivamente. En general si las apuestas en favor de una afirmación son de c a d, la probabilidad de ésta es $c/(c + d)$.

2.5 Desarrollo axiomático de la probabilidad

Para formalizar la definición de probabilidad, a través de un conjunto de axiomas, se repasarán brevemente los conceptos básicos de la teoría de conjuntos (o eventos), sobre los cuales se fundamenta la definición formal de probabilidad. Esta definición es tan general que permite incorporar las distintas interpretaciones de la probabilidad, mencionadas anteriormente.

La colección de todos los posibles resultados de un experimento aleatorio es importante en la definición de la probabilidad. Para definir esta colección considérense los siguientes experimentos: el número de reservaciones no canceladas para un vuelo, el número de llegadas a un servicio o la duración de un determinado componente. Todos son ejemplos de fenómenos impredecibles con un determinado número de posibles resultados. El número de reservaciones no canceladas puede ser cualquier entero positivo no mayor que el número de asientos del avión; el número de llegadas puede ser, teóricamente, cualquier entero positivo sin ningún límite, y la duración de un componente puede ser cualquier número real positivo. Lo anterior lleva, de manera inmediata, a la siguiente definición:

Definición 2.3 El conjunto de todos los posibles resultados de un experimento aleatorio recibe el nombre de *espacio muestral.*

El conjunto de todos los posibles resultados puede ser finito, infinito numerable o infinito no numerable. Por ejemplo, el número de reservaciones sin cancelar constituye un espacio muestral finito, dado que este número nunca excederá la capacidad del avión, que es finita. El número de llegadas al servicio constituye un espacio muestral infinito numerable, dado que es posible colocar los resultados en una correspondencia uno a uno con los enteros positivos, que constituyen un conjunto infinito pero numerable. La duración de una componente constituye un espacio muestral infinito innumerable, dado que ésta puede ser cualquier número real positivo. En este momento, es conveniente dar las siguientes definiciones.

Definición 2.4 Se dice que un espacio muestral es *discreto* si su resultado puede ponerse en una correspondencia uno a uno con el conjunto de los enteros positivos.

Definición 2.5 Se dice que un espacio muestral es *continuo* si sus resultados consisten de un intervalo de números reales.

Con respecto a los resultados de un espacio muestral, se puede estar particularmente interesado en un subconjunto de éstos. Por ejemplo, un gerente de cierta línea

aérea desea saber si el número de reservaciones sin cancelar es menor que cinco, o bien un comprador de baterías desea saber si éstas tendrán una operación normal mayor de 40 horas. De esta manera, se tiene la siguiente definición:

Definición 2.6 Un *evento* del espacio muestral es un grupo de resultados contenidos en éste, cuyos miembros tienen una característica común.

Por característica común debe entenderse que únicamente un grupo de resultados en particular satisface la característica y los restantes, contenidos en el espacio muestral, no. Se dice que ha ocurrido un evento si los resultados del experimento aleatorio incluyen a algunos de los que definen al evento. En este contexto, el espacio muestral, evento en sí mismo, puede entenderse como un *evento seguro*, puesto que se tiene un 100% de certidumbre de que ocurrirá un resultado del espacio muestral cuando el experimento se lleve a cabo. Para completar se dan las siguientes definiciones:

Definición 2.7 El evento que contiene a ningún resultado del espacio muestral recibe el nombre de *evento nulo* o *vacío.*

Deberán recordarse algunas definiciones de la teoría de eventos. Sean E_1 y E_2 cualesquiera dos eventos que se encuentren en un espacio muestral dado denotado por S.

Definición 2.8 El evento formado por todos los posibles resultados en E_1 o E_2 o en ambos, recibe el nombre de la *unión* de E_1 y E_2 y se denota por $E_1 \cup E_2$.

Definición 2.9 El evento formado por todos los resultados comunes tanto a E_1 como a E_2 recibe el nombre de *intersección de* E_1 y E_2 y se denota por $E_1 \cap E_2$.

Definición 2.10 Se dice que los eventos E_1 y E_2 son *mutuamente excluyentes* o *disjuntos* si no tienen resultados en común; en otras palabras $E_1 \cap E_2 = \emptyset \equiv$ evento vacío.

Definición 2.11 Si cualquier resultado de E_2 también es un resultado de E_1, se dice que el evento E_2 está *contenido* en E_1, y se denota por $E_2 \subset E_1$.

Definición 2.12 El *complemento* de un evento E con respecto al espacio muestral S, es aquel que contiene a todos los resultados de S que no se encuentran en E, y se denota por $\overline{E}$.

Las definiciones anteriores pueden demostrarse de manera gráfica mediante el uso de los diagramas de Venn, como se muestra en la figura 2.1.

Como ejemplo, considérese el experimento de lanzar un dado; el espacio muestral es S (1, 2, 3, 4, 5, 6). Se definen los eventos $E_1 = (2, 4, 6)$, $E_2 = (1, 3)$, y $E_3 = (2, 4)$. Es fácil verificar que $E_1 \cup E_2 = (1, 2, 3, 4, 6)$, $E_1 \cap E_3 = (2, 4)$, $E_1 \cap E_2 = \emptyset$, E_3 se encuentra completamente contenido en E_1 y $\overline{E}_2 = (2, 4, 5, 6)$.

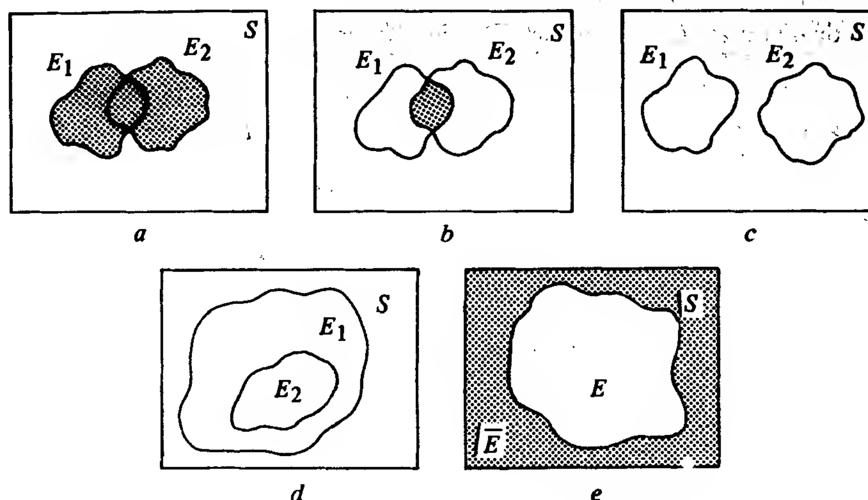


FIGURA 2.1 Diagramas de Venn que ilustran *a)* la unión de dos eventos; *b)* la intersección de dos eventos; *c)* eventos mutuamente excluyentes; *d)* un evento contenido en otro, y *e)* un evento y su complemento

La probabilidad es un número real que mide la posibilidad de que ocurra un resultado del espacio muestral, cuando el experimento se lleve a cabo. Por lo tanto, la probabilidad de un evento también es un número real que mide la posibilidad colectiva, de ocurrencia, de los resultados del evento cuando se lleve a efecto el experimento. A continuación se da la definición axiomática de la probabilidad.

Definición 2.13 Sean S cualquier espacio muestral y E cualquier evento de éste. Se llamará ***función de probabilidad*** sobre el espacio muestral S a $P(E)$ si satisface los siguientes axiomas:

1. $P(E) \geq 0$
2. $P(S) = 1$
3. Si, para los eventos $E_1, E_2, E_3, \ldots,$

   $E_i \cap E_j = \emptyset$ para toda $i \neq j$, entonces

   $P(E_1 \cup E_2 \cup \cdots) = P(E_1) + P(E_2) + \cdots.$

La razón de estos tres axiomas se convierte en aparente cuando, por ejemplo, se recuerda la interpretación de la probabilidad como una frecuencia relativa. Es decir, la probabilidad de un evento refleja la proporción de veces en que ocurrirá cuando el experimento se repita. Los axiomas también son evidentes para la interpretación

subjetiva de la probabilidad, dado que para ésta cualquier grado de creencia se convierte en una proporción. De ahí que las probabilidades exhiban las características de las proporciones, en las que la probabilidad es un número entre cero y uno, y dado que es forzoso que ocurra un resultado cuando se lleva a efecto un experimento, la probabilidad de S es uno. Además si no hay ningún resultado en común entre dos eventos E_1 y E_2, la probabilidad de que ocurra E_1 o E_2 es igual a la proporción de veces en que ocurre E_1 más la proporción de veces en que ocurra E_2.

En seguida se demostrarán algunas de las consecuencias de estos tres axiomas.

Teorema 2.1 $P(\emptyset) = 0$.

Demostración:

$$S \cup \emptyset = S \quad \text{y} \quad S \cap \emptyset = \emptyset.$$

Por el axioma 3,

$$P(S \cup \emptyset) = P(S) + P(\emptyset);$$

pero por el axioma 2, $P(S) = 1$, y de esta manera $P(\emptyset) = 0$.

Teorema 2.2 Para cualquier evento $E \subset S$, $0 \leqslant P(E) \leqslant 1$.

Demostración: Por el axioma 1, $P(E) \geqslant 0$; de aquí que sólo es necesario probar que $P(E) \leqslant 1$.

$$E \cup \overline{E} = S \quad \text{y} \quad E \cap \overline{E} = \emptyset.$$

Por los axiomas 2 y 3,

$$P(E \cup \overline{E}) = P(E) + P(\overline{E}) = P(S) = 1;$$

dado que $P(\overline{E}) \geqslant 0$, $P(E) \leqslant 1$.

El axioma 3 da la probabilidad de la unión de dos eventos disjuntos. Por otro esta porción de la suma de $P(A)$ y $P(B)$. El teorema se reduce al axioma 3 cuando la probabilidad de la unión de dos eventos que no son, necesariamente, disjuntos? Para dar respuesta a las preguntas anteriores se enuncia el siguiente resultado general, el que usualmente recibe el nombre de regla de adición de probabilidades.

Teorema 2.3 Sea S un espacio muestral que contiene a cualesquiera dos eventos A y B; entonces,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Aun cuando no se pretende dar aquí una demostración formal del teorema, éste es intuitivamente razonable. $P(A)$ y $P(B)$ reflejan el número de veces en que ocurrirán los resultados de A y B, respectivamente. Sin embargo, y teniendo en cuenta lo

anterior, los resultados comunes serán contados dos veces con la necesidad de restar esta porción de la suma de $P(A)$ y $P(B)$. El teorema se reduce al axioma 3 cuando los eventos son disjuntos.

Ejemplo 2.1 Un sistema contiene dos componentes A y B, y se conecta de manera que éste funciona si cualesquiera de las componentes funciona. Se sabe que la probabilidad de que A funcione es $P(A) = 0.9$ y la de B es $P(B) = 0.8$ y la probabilidad de ambos es $P(A \cap B) = 0.72$. Determinar la probabilidad de que el sistema funcione.

La probabilidad de que el sistema trabaje es igual a la probabilidad de la unión entre A y B; de esta manera,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= 0.9 + 0.8 - 0.72 = 0.98.$$

2.6 Probabilidades conjunta, marginal y condicional

En esta sección se examinan los conceptos de probabilidad conjunta, marginal y condicional, y se desarrolla la ley de multiplicación de probabilidades. Considérese un experimento en el que se elige aleatoriamente una persona adulta que viva en una ciudad con n personas adultas, y se anotan sus características con respecto a su hábitos de fumador y su sexo. Sea el espacio muestral la población de adultos de la ciudad, que se divide en los siguientes eventos disjuntos: fumador A_1 y no fumador A_2, hombre B_1 y mujer B_2. Los eventos en S pueden representarse como se muestra en la tabla 2.3.

Como ejemplo, nótese que n_{11} de los n adultos son hombres que fuman, por lo que son poseedores de los atributos A_1 y B_1. Supóngase que se desea determinar la probabilidad de ocurrencia simultánea de los eventos A_1 y B_2. Mediante el empleo de la interpretación de frecuencia relativa, puede argumentarse que, dado que exactamente n_{12} de los n adultos poseen ambos atributos, A_1 y B_2, la probabilidad es n_{12}/n. Esta última recibe el nombre de *probabilidad conjunta* puesto que se insiste en la probabilidad de resultados comunes a ambos eventos A_1 y B_2. Por lo tanto la probabilidad de los eventos A_i y B_j está dada por

$$P(A_i \cap B_j) = n_{ij}/n.$$

TABLA 2.3 Clasificación de n adultos mediante su sexo y hábitos de fumadores

	B_1	B_2
A_1	n_{11}	n_{12}
A_2	n_{21}	n_{22}

Supóngase que ahora el interés recae en determinar la probabilidad A_i, sin considerar cualquier otro evento B_j del espacio muestral S. Para especificar, supóngase que se necesita la probabilidad del evento A_2. Haciendo uso de nuevo de la interpretación de frecuencia relativa, el número total de personas no fumadoras (A_2) es $n_{21} + n_{22}$; de esta manera se tiene

$$P(A_2) = (n_{21} + n_{22})/n.$$

Este tipo de probabilidad se conoce como *marginal* porque para determinarla se ignoran una o más características del espacio muestral. De lo anterior se sigue que

$$P(A_i) = \sum_{j=1}^{2} n_{ij}/n,$$

pero dado que

$$P(A_i \cap B_j) = n_{ij}/n,$$

$$\mathrm{P}(A_i) = \sum_{j=1}^{2} P(A_i \cap B_j).$$

En otras palabras, la probabilidad marginal de un evento A_i es igual a la suma de las probabilidades conjuntas de A_i y B_j, donde la suma se efectúa sobre todos los eventos B_j. De manera similar la probabilidad marginal de B_j está dada por

$$P(B_j) = \sum_{i=1}^{2} P(A_i \cap B_j).$$

En este punto ya debe ser obvia la extensión para incluir más de dos eventos disjuntos.

Finalmente, supóngase que el interés recae en determinar la probabilidad de un evento A_i, dado que ha ocurrido el evento B_j. Por ejemplo, regresando a la tabla 2.3, supóngase que se ha elegido aleatoriamente una mujer adulta. (B_2) Ahora bien, ¿cuál es la probabilidad de que fume? Una vez más, el argumento descansa sobre la interpretación de frecuencia relativa. Sin embargo, una vez que el evento "mujer" ha ocurrido, éste reemplaza a S como el espacio muestral de interés. Por lo tanto, la probabilidad de tener un fumador (A_1) es el número de mujeres que fuman (n_{12}) entre el número total de estas ($n_{12} + n_{22}$). Por lo tanto

$$P(A_1|B_2) = n_{12}/(n_{12} + n_{22}),$$

donde la barra vertical se lee como "dado que" y separa al evento A_1, cuya probabilidad está condicionada a la previa ocurrencia del evento B_2. Ésta recibe el nombre de *probabilidad condicional* de A_1 dada la ocurrencia de B_2. En general, se tiene que

$$P(A_i|B_j) = n_{ij}/\sum_{i=1}^{2} n_{ij}, \tag{2.1}$$

y por simetría,

$$P(B_j|A_i) = n_{ij}/\sum_{j=1}^{2} n_{ij}. \tag{2.2}$$

Al dividir el numerador y denominador del miembro derecho de (2.1) por n, se tiene

$$P(A_i|B_j) = \frac{n_{ij}/n}{\sum_{i=1}^{2} n_{ij}/n},$$

pero

$$P(A_i \cap B_j) = n_{ij}/n$$

y

$$P(B_j) = \sum_{i=1}^{2} n_{ij}/n;$$

por lo tanto

$$P(A_i|B_j) = \frac{P(A_i \cap B_j)}{P(B_j)}, \quad P(B_j) > 0, \tag{2.3}$$

y de manera equivalente

$$P(B_j|A_i) = \frac{P(A_i \cap B_j)}{P(A_i)}, \quad P(A_i) > 0. \tag{2.4}$$

Para definir las probabilidades conjunta, marginal y condicional se ha empleado un ejemplo específico en el que el espacio muestral contiene únicamente un número finito de resultados. Sin embargo, las definiciones dadas aquí son completamente generales y pueden extenderse para incluir cualquier espacio muestral ya sea discreto o continuo. Con base en lo anterior se define de la siguiente manera.

Definición 2.14 Sean A y B cualesquiera dos eventos que se encuentran en un espacio muestral S de manera tal que $P(B) > 0$. La probabilidad condicional de A al ocurrir el evento B, es el cociente de la probabilidad conjunto de A y B con respecto a la probabilidad marginal de B; de esta manera se tiene

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0. \tag{2.5}$$

La relación entre (2.5) puede escribirse como un producto, lo que da como resultado la regla de multiplicación de probabilidades, dada por

$$P(A \cap B) = P(B)P(A|B). \tag{2.6}$$

Por simetría, la probabilidad condicional de B dada la ocurrencia de A, es

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \quad P(A) > 0.$$

De esta manera se tiene

$$P(A \cap B) = P(A)P(B|A)$$

que es otra versión de la regla de multiplicación, la que implica que

$$P(A)P(B|A) = P(B)P(A|B). \tag{2.7}$$

La definición 2.14 puede extenderse para incluir cualquier número de eventos que se encuentren en el espacio muestral. Por ejemplo, puede demostrarse que para tres eventos A, B y C

$$P(A|B \cap C) = \frac{P(A \cap B \cap C)}{P(B \cap C)}, \quad P(B \cap C) > 0 \tag{2.8}$$

y

$$P(A \cap B|C) = \frac{P(A \cap B \cap C)}{P(C)}, \quad P(C) > 0. \tag{2.9}$$

Los siguientes ejemplos ilustrarán los conceptos presentados en esta sección.

Ejemplo 2.2 A los habitantes de una gran ciudad se les hizo una encuesta con el propósito de determinar el número de lectores de *Time* y *Newsweek*. Los resultados de la encuesta fueron los siguientes: 20% de los habitantes leen el *Time*, el 16% lee el *Newsweek* y un 1% lee ambos semanarios. Si se selecciona al azar a un lector de *Time*, ¿cuál es la probabilidad de que también lea el *Newsweek?*

Sean A y B los eventos que representan el número de lectores del *Time* y *Newsweek* respectivamente; dado que $P(A) = 0.2$, $P(B) = 0.16$ y $P(A \cap B) = 0.01$,

$$P(\mathrm{B}|\mathrm{A}) = 0.01/0.2 = 0.05.$$

Por otra parte, también puede determinarse la probabilidad de que un lector del *Newsweek* lea también el *Time*; esto es

$$P(A|B) = 0.01/0.16 = 0.0625,$$

y se verifica la relación $P(A)P(B|A) = P(B)P(A|B)$, o (0.2)(0.05) (0.16)(0.0625).

Ejemplo 2.3 Muchas instituciones bancarias emplean modelos computarizados de crédito con el propósito de dar un determinado puntaje a todas las solicitudes de préstamo. Este puntaje se emplea como una ayuda para decidir cuándo se otorga el préstamo. Supóngase que el 3% de todos los préstamos que se otorgan presentan problemas por incumplimiento de pago y que los modelos de crédito son precisos en

un 80% al predecir menos créditos. Si el 85% de todas las solicitudes reciben puntuaciones favorables por los modelos computarizados y se les otorga el préstamo, determinar la probabilidad de que una solicitud que recibe una puntuación favorable y a la que se le otorga el préstamo, no presente ningún problema para el pago de éste.

Sea A el evento incumplimiento de pago y B la puntuación favorable. Del enunciado del problema se tiene que $P(A) = 0.03$, $P(B) = 0.85$ y $P(B|\overline{A}) = 0.8$, en donde $\overline{A}$ es el complemento de A, es decir, el evento cumplimiento de pago. Lo que se busca es la probabilidad condicional de que no exista ningún problema en el pago del préstamo, dado que la solicitud obtuvo una puntuación favorable, o $P(\overline{A}|B)$. Usando la relación (2.7), se tiene

$$P(B)P(\overline{A}|B) = P(\overline{A})P(B|\overline{A}),$$

o

$$P(\overline{A}|B) = \frac{P(\overline{A})P(B|\overline{A})}{P(B)},$$

y dado que $P(\overline{A}) = 0.97$, la probabilidad deseada es $P(\overline{A}|B) = 0.9129$.

Ejemplo 2.4 Una planta recibe reguladores de voltaje de dos diferentes proveedores, B_1 y B_2; el 75% de los reguladores se compra a B_1 y el resto a B_2. El porcentaje de reguladores defectuosos que se reciben de B_1 es 8% y el de B_2 es 10%. Determinar la probabilidad de que funcione un regulador de voltaje de acuerdo con las especificaciones (es decir, el regulador no está defectuoso).

Sea A el evento el regulador de voltaje es no defectuoso. Es claro que ningún regulador de voltaje puede ser vendido tanto por B_1 como por B_2; por lo tanto B_1 y B_2 son disjuntos. Esto da como resultado

$$P(A) = P(A \cap B_1) + P(A \cap B_2),$$

pero

$$P(A \cap B_1) = P(B_1)P(A|B_1)$$

y

$$P(A \cap B_2) = P(B_2)P(A|B_2),$$

en donde se conocen $P(B_1) = 0.75$, $P(B_2) = 0.25$, $P(A|B_1) = 0.92$, y $P(A|B_2) = 0.9$; sustituyendo

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2)$$

$$= (0.75)(0.92) + (0.25)(0.90) = 0.915.$$

Nótese que en el ejemplo 2.4 se tienen únicamente dos proveedores, B_1 y B_2. En general, si existen n alternativas disjuntas B_1, B_2 ... B_n, la probabilidad total de un

$B_1, B_2, \ldots, B_n,$

resultado final, por ejemplo A, está dada por

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i). \tag{2.10}$$

2.7 Eventos estadísticamente independientes

Al considerar la probabilidad condicional de algún evento A, dada la ocurrencia de otro evento B, siempre se implica que las probabilidades de A y B son de alguna manera dependientes entre sí. En otras palabras, la información con respecto a la ocurrencia de B afectará la probabilidad de A. Supóngase que la ocurrencia de B no tiene ningún efecto sobre la probabilidad de A, en el sentido de que la probabilidad condicional $P(A|B)$ es igual a la probabilidad marginal $P(A)$, aun a pesar de que haya ocurrido el evento B. Esta situación origina un concepto muy importante que se conoce como independencia estadística.

Definición 2.15 Sean A y B dos eventos cualesquiera de un espacio muestral S. Se dice que el evento A es *estadísticamente independiente* del evento B si $P(A|B) = P(A)$.

Algunas consecuencias de la definición 2.15 se convierten en evidentes en este momento, dado que

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

si A es independiente de B,

$$P(A|B) = P(A) = \frac{P(A \cap B)}{P(B)}$$

o

$$P(A \cap B) = P(A)P(B).$$

Además, puesto que

$$P(A \cap B) = P(A)P(B|A),$$

entonces

$$P(A)P(B) = P(A)P(B|A)$$

o

$$P(B) = P(B|A).$$

Por lo tanto, puede concluirse que si un evento A es estadísticamente independiente

de B, entonces el evento B es independiente de A y se verifican las tres relaciones siguientes:

1. $P(A|B) = P(A)$,
2. $P(B|A) = P(B)$, y
3. $P(A \cap B) = P(A)P(B)$.

Con la siguiente definición se extenderá el concepto de independencia estadística.

Definición 2.16 Los eventos $A_1, A_2 \dots A_k$ de un espacio muestral S son estadísticamente independientes si y sólo si la probabilidad conjunta de cualquier 2, 3 ... k de ellos es igual al producto de sus respectivas probabilidades marginales.

De esta manera, los eventos A, B y C son estadísticamente independientes, sí y sólo si

1. $P(A \cap B) = P(A)P(B)$,
2. $P(A \cap C) = P(A)P(C)$,
3. $P(B \cap C) = P(B)P(C)$, y
4. $P(A \cap B \cap C) = P(A)P(B)P(C)$

Ejemplo 2.5 Un sistema contiene cinco componentes que se encuentran conectadas entre sí como se muestra en la figura 2.2, donde las probabilidades indican la seguridad de que la componente funcione adecuadamente. Si se supone que el funcionamiento de una componente en particular es independiente del de las demás, ¿cuál es la probabilidad de que el sistema trabaje?

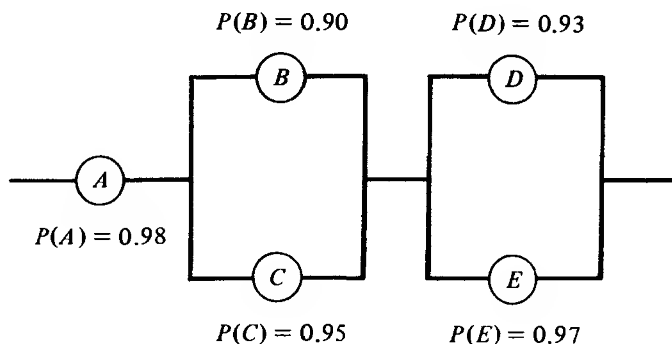


FIGURA 2.2 Configuración de un sistema con cinco componentes

Establecida la suposición de independencia, el sistema puede trabajar si las componentes A y B y/o C, y D y/o E lo hacen. De esta manera, la probabilidad de que el sistema trabaje, $P(F)$, puede expresarse como

$$P(F) = P(A)P(B \cup C)P(D \cup E);$$

pero nótese que $P(B \cup C) = 1 - P(\overline{B})P(\overline{C})$ y $P(D \cup E) = 1 - P(\overline{D})P(\overline{E})$, porque, por ejemplo $P(\overline{B})P(\overline{C})$ es la probabilidad de que no trabaje la componente B y tampoco la C. Por lo tanto,

$$P(F) = (0.98)(0.995)(0.9979) = 0.973.$$

2.8 El teorema de Bayes

Recuérdese el ejemplo 2.4. Supóngase que cuando se reciben los reguladores de voltaje se almacenan de manera tal que no puede distinguirse el proveedor. Además, supóngase que se desea determinar la probabilidad de que un regulador en particular fue vendido por el proveedor B_2 cuando se sabe que funciona de acuerdo con las especificaciones. En este caso se busca la probabilidad condicional de B_2 dada la ocurrencia del evento A. Por lo tanto

$$P(B_2|A) = \frac{P(B_2 \cap A)}{P(A)},$$

pero

$$P(B_2 \cap A) = P(B_2)P(A|B_2)$$

y

$$P(B_2|A) = \frac{P(B_2)P(A|B_2)}{P(A)};$$

así que,

$$P(B_2|A) = \frac{(0.25)(0.9)}{0.915} = 0.2459.$$

Se puede generalizar el método empleado para resolver este problema, con el fin de originar el teorema de Bayes.

Teorema 2.4 Si $B_1, B_2, \ldots, B_n$ son n eventos mutuamente excluyentes, de los cuales uno debe ocurrir, es decir $\sum_{i=1}^{n} P(B_i) = 1$, entonces

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}, \quad j = 1, 2, \ldots, n. \tag{2.11}$$

La expresión dada por (2.11) fue desarrollada por el reverendo Thomas Bayes (1702-1761) y se conoce como teorema de Bayes. A primera vista no es más que una aplicación de las probabilidades condicionales. Sin embargo, ha sido clave en el

desarrollo de la inferencia estadística bayesiana en la que se emplea la interpretación subjetiva de la probabilidad. Tal como se indicó en el capítulo uno, la inferencia bayesiana no se tratará con detalle en este libro, Sin embargo, se considerarán algunas cuestiones bayesianas de vez en cuando, de manera que el lector pueda obtener una mejor perspectiva de la inferencia estadística. Los siguientes son ejemplos del análisis bayesiano.

Supóngase que un investigador conduce un experimento en el que sabe que el resultado de interés estará afectado por cualquiera de las n alternativas B_1, B_2 ... B_n que predomine. A pesar de que no está seguro cuál de todas las alternativas predominará, posee cierta información con base en la cual está dispuesto a formular un juicio subjetivo para las probabilidades de ocurrencia de las n alternativas. De esta forma, asigna probabilidades $P(B_1)$, $P(B_2)$ ... $P(B_n)$ para las n alternativas antes de obtener cualquier evidencia experimental. Dado que estas probabilidades reflejan el juicio o grado de creencia del investigador con respecto a las ocurrencias de B_1, B_2 ... B_n, *antes de que éstas se presenten se conocen como probabilidades a priori.* Con ello el investigador obtendrá una evidencia experimental a partir de un conjunto de datos que se denota por A, y se observa bajo una alternativa específica B_j. En este momento se pueden calcular las probabilidades condicionales $P(A|B_j)$. Éstas permitirán la determinación de la probabilidad B_j dada la evidencia experimental A, mediante el empleo del teorema de Bayes. Las probabilidades condicionales $P(B_j|A)$, $j = 1, 2, \ldots, n$ se conocen como probabilidades *a posteriori* porque se determinan una vez obtenida la evidencia experimental. Por lo tanto, las probabilidades $P(B_j|A)$ reflejan el grado de creencia corregido con respecto a las alternativas B_1, B_2, ... B_n después de obtener los datos experimentales.

Ejemplo 2.6 Durante los últimos años se ha escrito mucho sobre la posible relación entre el fumar y el cáncer pulmonar. Supóngase que en un centro médico, de todos los fumadores de quienes se sospecha que tenían cáncer pulmonar, el 90% lo tenía mientras que únicamente el 5% de los no fumadores lo padecía. Si la proporción de fumadores es de 0.45, ¿cuál es la probabilidad de que un paciente con cáncer pulmonar, seleccionado al azar, sea fumador?

Sean B_1 y B_2 los eventos "el paciente es fumador" y "el paciente es no fumador" respectivamente, y sea A el evento "el paciente tiene cáncer pulmonar". B_1 y B_2 son las alternativas que pueden predominar. Se supone que las probabilidades *a priori*, para estas dos alternativas, son 0.45 y 0.55 respectivamente. Si un paciente tiene o no cáncer pulmonar puede estar afectado por cualquiera de las dos alternativas que predominen y que constituyen la evidencia experimental. Se sabe que $P(A|B_1) = 0.9$ y $P(A|B_2) = 0.05$. Se desea determinar la probabilidad *a posteriori* de seleccionar un fumador, puesto que el paciente tiene cáncer, o $P(B_1|A)$.

Del teorema de Bayes se tiene

$$P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2)}$$

$$= \frac{(0.45)(0.9)}{(0.45)(0.9) + (0.55)(0.05)}$$

$$= 0.9364.$$

La probabilidad de que un paciente con cáncer pulmonar, seleccionado aleatoriamente sea fumador, es de 0.9364.

Ejemplo 2.7 **Una compañía estudia la comercialización de un nuevo producto. El presidente de la compañía desea que el producto sea superior al de su más cercano competidor. Con base en una evaluación preliminar que realizó el personal clave, se decide asignar una posibilidad del 50% de que el producto sea superior al ofrecido por el competidor, 30% de que tenga la misma calidad y un 20% de que sea inferior. Un estudio de mercado sobre el producto concluye que éste es superior al del competidor. Con base en la experiencia sobre los resultados de las encuestas, se determina que si el producto realmente es superior, la probabilidad de que la encuesta alcance la misma conclusión es 0.7. Si el producto tiene la misma calidad que el del competidor, la probabilidad de que la encuesta dé como resultado un producto superior es 0.4. Si el producto es inferior, la probabilidad de que la encuesta indique un producto superior es de 0.2. Dado el resultado de la encuesta, ¿cuál es la probabilidad, corregida, de obtener un producto superior?**

Este es un ejemplo en el que ilustra cómo una organización puede actualizar y revisar las probabilidades iniciales al tener disponible nueva información. Sean B_1, B_2 y B_3 los eventos el producto es superior, tiene la misma calidad y es inferior al del competidor, respectivamente. Las probabilidades *a priori* correspondientes son 0.5, 0.3 y 0.2. Sea A el evento "la encuesta revelará un producto superior". Las probabilidades condicionales que involucran una evidencia experimental son $P(A|B_1) = 0.7$, $P(A|B_2) = 0.4$ y $P(A|B_3) = 0.2$. La probabilidad *a posteriori* $P(B_1|A)$ deseada es:

$$P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)}$$

$$= 0.6863.$$

2.9 Permutaciones y combinaciones

Para calcular las probabilidades de varios eventos es necesario contar el número de resultados posibles de un experimento, o contar el número de resultados que son favorables a un evento dado. El proceso de conteo puede simplificarse mediante el empleo de dos técnicas de conteo denominadas permutaciones y combinaciones.

Una *permutación* es un arreglo en un orden particular, de los objetos que forman un conjunto. Por ejemplo, considere las diferentes formas en que pueden situarse las letras *a, b* y *c*. Para la primera posición puede elegirse a cualquiera de las tres letras; para la segunda se puede escoger a cualquiera de las dos restantes y para la tercera debe seleccionarse la letra que no se utilizó. Así existen $3 \times 2 \times 1 = 6$ maneras en las que pueden arreglarse tres letras. Los seis arreglos o permutaciones son:

abc, acb, bac, bca, cab, cba.

empleando el mismo razonamiento, el número total de maneras en que pueden arreglarse las letras *a, b, c* y *d* es $4 \times 3 \times 2 \times 1 = 24$. En general, el número de permutaciones de n objetos diferentes es:

$$n(n-1)(n-2)\cdots(2)(1). \tag{2.12}$$

El producto de un entero positivo por todos los que le preceden se denota por $n!$ y see lee "n factorial". Por ejemplo, $2! = 2 \times 1 = 2$, $3! = 3 \times 2 \times 1 = 6$, $4! = 4 \times 3 \times 2 \times 1 = 24$, etc. Nótese que de (2.12) se tiene:

$$n(n-1)! = n!$$

o

$$(n-1)! = n!/n.$$

De esta manera, cuando $n = 1$, se define a $0! = 1$.

En este punto se examinarán las permutaciones de n objetos, si únicamente $r \leq n$ de éstos se emplean en cualquier ordenamiento. Igualmente, para la primera posición se puede seleccionar cualquiera de los n objetos, para la segunda uno de los restantes $n - 1$, y se continúa el procedimiento hasta la r-ésima posición. En este momento se han empleado $r - 1$ objetos, quedando $n - (r - 1)$, a partir de los cuales se hace la selección. Por lo tanto, el número de permutaciones de n objetos si se toma r a la vez es:

$$\begin{aligned} P(n, r)^* &= n(n-1)(n-2)\cdots(n-r+1) \\ &= \frac{n(n-1)(n-2)\cdots(n-r+1)(n-r)!}{(n-r)!} \\ &= \frac{n!}{(n-r)!}. \end{aligned} \tag{2.13}$$

Nótese que si $r = n$, (2.13) se reduce al resultado anterior $P(n, n) = n!$, o el número de permutaciones de n objetos, tomando n a la vez, es $n!$.

Ejemplo 2.8 En muchos Estados de la Unión Americana, las placas de los automóviles, se identifican por tres letras y tres números. ¿Cuál es el número total si ninguna letra de placas posible puede usarse más de una ocasión en la misma placa? ¿Cuál es el número total sin esta restricción?

Con la restricción, el número de permutaciones que puede obtenerse con las 26 letras del alfabeto, tomadas tres a la vez, es:

$$P(26, 3) = \frac{26!}{23!} = \frac{26 \times 25 \times 24 \times 23!}{23!} = 15\,600.$$

* Esta es una de las muchas formas de denotar el número de permutaciones de n objetos tomando r a la vez. Otros símbolos empleados son ${}_nP_r$, P^n_r, $P_{n,r}$ y $(n)_r$.

Dado que a cada uno de los 15 600 arreglos de tres letras se les puede asignar 1 000 diferentes números de tres dígitos (000-999), el número total de placas es de 15 600 000. Sin la restricción, que es la práctica usual, las seis posiciones en una placa de automóvil pueden ocuparse de la siguiente forma: cada una de las tres primeras posiciones puede ocuparse de 26 maneras diferentes, mientras que cada una de las tres posiciones restantes puede ocuparse en una de diez formas posibles; dado que existen 26 letras y diez números, respectivamente. De esta manera el número total de placas de automóvil es $26 \times 26 \times 26 \times 10 \times 10 \times 10 = 17\,576\,000$.

Una *combinación* de los objetos de un conjunto es una selección de éstos sin importar el orden. Se entenderá por el número de combinaciones de r objetos tomados de un conjunto que contiene a n de éstos, al número total de selecciones distintas en las que cada una de éstas contiene r objetos. La diferencia entre una permutación y una combinación es que en la primera el interés se centra en contar todas las posibles selecciones y todos los arreglos de éstas, mientras que en la segunda el interés sólo recae en contar el número de selecciones diferentes. De esta manera *abc* y *acd* son diferentes combinaciones de tres letras, mientras que *acd* y *adc* son distintas permutaciones de la misma combinación. Puede obtenerse el número de combinaciones de n objetos tomando r a la vez (denotada por $\binom{n}{r}$) y que se lee "n combinación r"), dividiendo el correspondiente número de permutaciones por $r!$ dado que en cada combinación existen $r!$ permutaciones. Por lo tanto:

$$\binom{n}{r}^{*} = P(n, r)/r!$$

$$= \frac{n!}{(n-r)!\,r!}. \qquad (2.14)$$

De (2.14) puede notarse que:

$$\binom{n}{n} = \frac{n!}{(n-n)!\,n!} = 1;$$

$$\binom{n}{0} = \frac{n!}{(n-0)!\,0!} = 1;$$

$$\binom{n}{n-1} = \frac{n!}{[n-(n-1)]!(n-1)!} = n;$$

y

$$\binom{n}{n-r} = \frac{n!}{[n-(n-r)]!(n-r)!} = \binom{n}{r}.$$

Dos ejemplos específicos son:

$$\binom{5}{2} = \frac{5!}{(5-2)!\,2!} = \frac{5 \times 4 \times 3!}{3!2!} = 10,$$

* Otros símbolos comúnmente empleados para denotar el número de combinaciones de n objetos, tomando r a la vez, son $C(n,r)$, ${}_nC_r$, C_r^n, y $C_{n,r}$.

y

$$\binom{10}{8} = \binom{10}{2} = \frac{10!}{(10-2)!\,2!} = \frac{10 \times 9 \times 8!}{8!\,2!} = 45.$$

Ejemplo 2.9 Supóngase que van a enviarse cinco jueces federales a cierto Estado. El jefe del senado estatal envía al presidente una lista que contiene los nombres de diez hombres y cuatro mujeres. Si el presidente decide que de los cinco jueces tres deben ser hombres y dos mujeres ¿de cuántas maneras puede lograrse lo anterior, empleando a los candidatos de la lista?

El número de maneras distintas en que pueden seleccionarse tres hombres de entre diez es:

$$\binom{10}{3} = \frac{10 \times 9 \times 8 \times 7!}{7!\,3!} = 120.$$

Asimismo, el número de maneras en que pueden seleccionarse dos mujeres de entre cuatro es:

$$\binom{4}{2} = \frac{4 \times 3 \times 2!}{2!\,2!} = 6.$$

Puesto que el número de maneras en que pueden seleccionarse tres hombres de entre diez es 120, y el de dos mujeres de entre cuatro es seis, el número de maneras en que ambos eventos pueden ocurrir es:

$$\binom{10}{3}\binom{4}{2} = 720.$$

Referencias

1. P. G. Hoel, *Introduction to mathematical statistics*, 4th ed., Wiley, New York, 1971.
2. A. M. Mood and F. A. Graybill, *Introduction to the theory of statistics*, 2nd ed., McGraw-Hill, New York, 1963.

Ejercicios

2.1. Los empleados de la compañía New Horizons se encuentran separados en tres divisiones: administración, operación de planta y ventas. La siguiente tabla indica el número de empleados en cada división clasificados por sexo:

	Mujer (M)	*Hombre (H)*	*Totales*
Administración (A)	20	30	50
Operación de planta (*O*)	60	140	200
Ventas (*V*)	100	50	150
Totales	180	220	400

a) Usar un diagrama de Venn para ilustrar los eventos O y M para todos los empleados de la compañía. ¿Son mutuamente excluyentes?

b) Si se elige aleatoriamente un empleado:

1. ¿Cuál es la probabilidad de que sea mujer?
2. ¿Cuál es la probabilidd de que trabaje en ventas?
3. ¿Cuál es la probabilidad de que sea hombre y trabaje en la división de administración?
4. ¿Cuál es la probabilidad de que trabaje en la división de operación de planta, si es mujer?
5. ¿Cuál es la probabilidad de que sea mujer si trabaja en la división de operación de planta?

c) ¿Son los eventos V y H estadísticamente independientes?

d) ¿Son los eventos A y M estadísticamente independientes?

e) Determinar las siguientes probabilidades:

1. $P(A \cup M)$
2. $P(A \cup \overline{M})$
3. $P(O \cap F)$
4. $P(M|A)$

2.2. Con la definición 2.14 demuéstrese que para cualesquiera dos eventos, A y B, $P(A|B) + P(\overline{A}|B) = 1$, con tal de que $P(B) \neq 0$.

2.3. Sean A y B dos eventos cualquiera de S. Si A y B son mutuamente excluyentes, muéstrese que no pueden ser independientes. Dedúzcase cuándo dos eventos independientes son, también, mutuamente excluyentes.

2.4. Sean A y B dos eventos cualquiera de S. Empléese un diagrama de Venn para demostrar que $P(A \cap \overline{B}) = P(A) - P(A \cap B)$.

2.5. Una familia tiene tres hijos. Determinar todas las posibles permutaciones, con respecto al sexo de los hijos. Bajo suposiciones adecuadas, ¿cuál es la probabilidad de que, exactamente, dos de los hijos tengan el mismo sexo?, ¿cuál es la probabilidad de tener un varón y dos mujeres?, ¿cuál es la probabilidad de tener tres hijos del mismo sexo?

2.6. Se extraen, sin reemplazo, dos cartas de una baraja. ¿Cuál es la probabilidad de que ambas sean ases?

2.7. Se lanza una moneda diez veces y en todos los lanzamientos el resultado es cara. ¿Cuál es la probabilidad de este evento?, ¿cuál es la probabilidad de que en el decimoprimero lanzamiento el resultado sea cruz?

2.8. Una agencia automotriz recibe un embarque de 20 automóviles nuevos. Entre éstos, dos tienen defectos. La agencia decide seleccionar, aleatoriamente, dos automóviles de entre los 20 y aceptar el embarque si ninguno de los dos vehículos seleccionados tiene defectos. ¿Cuál es la probabilidad de aceptar el embarque?

2.9. Se lanza una moneda con una probabilidad de 2/3 que el resultado sea cara. Si aparece una cara, se extrae una pelota, aleatoriamente, de una urna que contiene dos pelotas rojas y tres verdes. Si el resultado es cruz se extrae una pelota, de otra urna, que contiene dos rojas y dos verdes. ¿Cuál es la probabilidad de extraer una pelota roja?

2.10. De entre 20 tanques de combustible fabricados para el transbordador espacial, tres se encuentran defectuosos. Si se seleccionan aleatoriamente cuatro tanques:

a) ¿Cuál es la probabilidad de que ninguno de los tanques se encuentre defectuoso?

b) ¿Cuál es la probabilidad de que uno de los tanques tenga defectos?

2.11. La probabilidad de que cierto componente eléctrico funcione es de 0.9. Un aparato contiene dos de éstos componentes. El aparato funcionará mientras lo haga, por lo menos, uno de los componentes.

a) Sin importar cuál de los dos componentes funcione o no, ¿cuáles son los posibles resultados y sus respectivas probabilidades? (Puede suponerse independiencia en la operación entre los componentes.)

b) ¿Cuál es la probabilidad de que el aparato funcione?

2.12. Un sistema contiene tres componentes *A*, *B* y *C*. Éstos pueden conectarse en una, cualquiera, de las cuatro configuraciones mostradas en la figura 2.3. Si los tres componentes operan de manera independiente y si la probabilidad de que uno, cualquiera de ellos, esté funcionando es de 0.95, determinar la probabilidad de que el sistema funcione para cada una de las cuatro configuraciones.

2.13. Una forma de incrementar la probabilidad de operación de un sistema (conocida como la confiabilidad del sistema), es mediante la introducción de una copia de los componentes en una configuración paralela, como se ilustra en la segunda parte de la figura 2.3. Supóngase que la Nasa desea una probabilidad no menor de 0.999 99, de que el transbordador espacial entre en órbita alrededor de la tierra, con éxito. ¿Cuántos motores cohete deben configurarse en paralelo para alcanzar esta confiabilidad de operación si se sabe que la probabilidad de que uno, cualquiera, de los motores funcione adecuadamente es de 0.95? Supóngase que los motores funcionan de manera independiente entre sí.

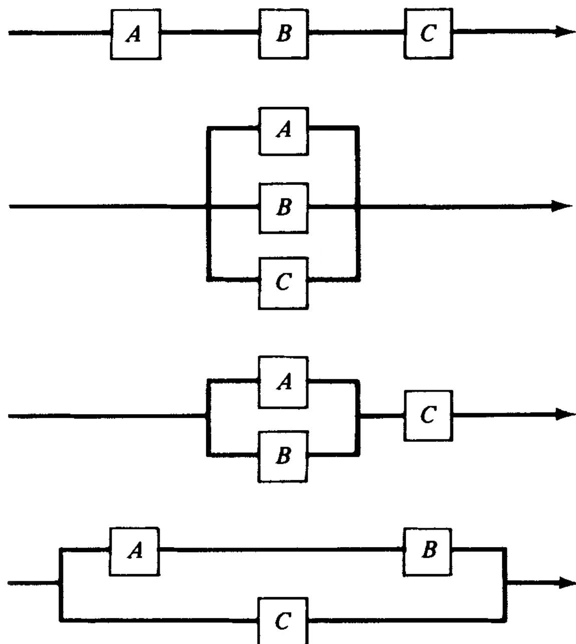


FIGURA 2.3 Cuatro configuraciones para tres componentes

2.14. Supóngase que la probabilidad de que los Potros de Baltimore ganen el campeonato de la Conferencia Americana es de 0.25, y la probabilidad de que lo obtengan los Cargadores de San Diego es de 0.20. Además, la probabilidad de que el campeón de la Conferencia Americana gane el Super Tazón es 0.45, 0.55 o 0.35, dependiendo de si los Potros, los Cargadores o algún otro equipo gana el campeonato.

a) ¿Cuál es la probabilidad de que un equipo de la Conferencia Americana gane el Super Tazón?

b) Si un equipo de la Conferencia Americana gana el Super Tazón, ¿cuál es la probabilidad de que los Potros de Baltimore ganen el título de su Conferencia?

2.15. El 5% de las unidades producidas en una fábrica se encuentran defectuosas cuando el proceso de fabricación se encuentra bajo control. Si el proceso se encuentra fuera de control, se produce un 30% de unidades defectuosas. La probabilidad marginal de que el proceso se encuentre bajo control es de 0.92. Si se escoge aleatoriamente una unidad y se encuentra que es defectuosa, ¿cuál es la probabilidad de que el proceso se encuentre bajo control?

2.16. Una planta armadora recibe microcircuitos provenientes de tres distintos fabricantes B_1, B_2 y B_3. El 50% del total se compra a B_1 mientras que a B_2 y B_3 se les compra un 25% a cada uno. El porcentaje de circuitos defectuosos para B_1, B_2 y B_3 es 5, 10 y 12% respectivamente. Si los circuitos se almacenan en la planta sin importar quién fue el proveedor:

a) Determinar la probabilidad de que una unidad armada en la planta contenga un circuito defectuoso.

b) Si un circuito no está defectuoso, ¿cuál es la probabilidad de que haya sido vendido por el proveedor B_2?

2.17. Un inversionista está pensando en comprar un número muy grande de acciones de una compañía. La cotización de las acciones en la bolsa, durante los seis meses anteriores, es de gran interés para el inversionista. Con base en esta información, se observa que la cotización se relaciona con el producto nacional bruto. Si el PNB aumenta, la probabilidad de que el valor de las acciones aumente es de 0.8. Si el PNB es el mismo, la probabilidad de que las acciones aumenten su valor es de 0.2. Si el PNB disminuye, la probabilidad es de sólo 0.1. Si para los siguientes seis meses se asignan las probabilidades 0.4, 0.3 y 0.3 a los eventos, el PNB aumenta, es el mismo y disminuye, respectivamente, determinar la probabilidad de que las acciones aumenten su valor en los próximos seis meses.

2.18. Con base en varios estudios una compañía ha clasificado, de acuerdo con la posibilidad de descubrir petróleo, las formaciones geológicas en tres tipos. La compañía pretende perforar un pozo en un determinado sitio, al que se le asignan las probabilidades de 0.35, 0.40 y 0.25 para los tres tipos de formaciones respectivamente. De acuerdo con la experiencia, se sabe que el petróleo se encuentra en un 40% de formaciones del tipo I, en un 20% de formaciones del tipo II y en un 30% de formaciones del tipo III. Si la compañía no descubre petróleo en ese lugar, determínese la probabilidad de que exista una formación del tipo II.

CAPÍTULO TRES

Variables aleatorias y distribuciones de probabilidad

3.1 El concepto de variable aleatoria

En el capítulo dos se examinaron los conceptos básicos de probabilidad con respecto a eventos que se encuentran en un espacio muestral. Los experimentos se conciben de manera que los resultados del espacio muestral son cualitativos o cuantitativos. Como ejemplos de resultados cualitativos se tienen: *a)* el lanzamiento de una moneda es "cara" o "cruz"; *b)* un producto manufacturado en una fábrica puede ser "defectuoso" o "no defectuoso", o *c)* una persona en particular puede preferir la loción X sobre la loción Y. Puede ser útil la cuantificación de los resultados cualitativos de un espacio muestral y, mediante el empleo de medidas numéricas, estudiar su comportamiento aleatorio. El concepto de variable aleatoria proporciona un medio para relacionar cualquier resultado con una medida cuantitativa.

Definición 3.1 Sea S un espacio muestral sobre el que se encuentra definida una función de probabilidad. Sea X una función de valor real definida sobre S, de manera que transforme los resultados de S en puntos sobre la recta de los reales. Se dice entonces que X es una *variable aleatoria.*

Se dice que X es "aleatoria" porque involucra la probabilidad de los resultados del espacio muestral, y X es una función definida sobre el espacio muestral, de manera que transforma todos los posibles resultados del espacio muestral en cantidades numéricas.

Par ilustrar la noción de variable aleatoria, considérese el lanzamiento de una moneda. El espacio muestral está constituido por dos posibles resultados, "cara" y "cruz". Sea $X(\text{cruz}) = 0$ y $X(\text{cara}) = 1$; de esta manera se han transformado los dos posibles resultados del espacio muestral en puntos sobre la recta de los reales. Por $P(X = 0)$ se entenderá la probabilidad de que la variable aleatoria tome el valor cero o, de manera equivalente, la probabilidad de que caiga cruz cuando se lance la moneda. Como ejemplo adicional, considérese el lanzamiento de dos dados

indistinguibles y los 36 posibles resultados, como se muestra en la tabla 2.1. Se define como variable aleatoria X a la suma de los valores de las dos caras de los dados. La tabla 3.1 relaciona los 36 resultados con los valores correspondientes de la variable aleatoria X y sus probabilidades. La naturaleza probabilística de la variable aleatoria X, la suma de las dos caras, puede observarse el graficar cada valor de X contra su probabilidad como se muestra en la figura 3.1.

Para cada uno de los ejemplos anteriores, el número de posibles valores de la variable aleatoria es finito. Sin embargo, se pueden definir variables aleatorias cuyos valores sean contables o no. Ya que una variable aleatoria es una caracterización cuantitativa de los resultados de un espacio muestral, ésta posee intrínsecamente la naturaleza discreta o continua de este espacio.

Definición 3.2 Se dice que una variable aleatoria X es *discreta* si el número de valores que puede tomar es contable (ya sea finito o infinito), y si éstos pueden arreglarse en una secuencia que corresponde con los enteros positivos.

Definición 3.3 Se dice que una variable aleatoria X es *continua* si sus valores consisten en uno o más intervalos de la recta de los reales.

3.2 Distribuciones de probabilidad de variables aleatorias discretas

En esta sección se considerará el concepto de distribución de probabilidad de una variable aleatoria. En la figura 3.1 se muestra la gráfica de los valores correspondientes a la variable aleatoria que respresenta la suma de las caras de los dos dados, cuando éstos se tiran. En general, una variable aleatoria discreta X representa los resultados de un espacio muestral en forma tal que por $P(X = x)$ se entenderá la probabilidad de que X tome el valor de x. De esta forma, al considerar los valores de una variable aleatoria es posible desarrollar una función matemática que asigne una probabilidad a cada *realización* x de la variable aleatoria X. Esta función recibe el nombre de *fun-*

TABLA 3.1 Correspondencia entre los resultados del lanzamiento de un par de dados y la variable aleatoria que representa la suma de las caras

Resultado	*Valor de la variable aleatoria*	*Número de ocurrencias*	*Probabilidad*
(1,1)	2	1	1/36
(1,2), (2,1)	3	2	2/36
(1,3), (2,2), (3,1)	4	3	3/36
(1,4), (2,3), (3,2), (4,1)	5	4	4/36
(1,5), (2,4), (3,3), (4,2), (5,1)	6	5	5/36
(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)	7	6	6/36
(2,6), (3,5), (4,4), (5,3), (6,2)	8	5	5/36
(3,6), (4,5), (5,4), (6,3)	9	4	4/36
(4,6), (5,5), (6,4)	10	3	3/36
(5,6), (6,5)	11	2	2/36
(6,6)	12	1	1/36

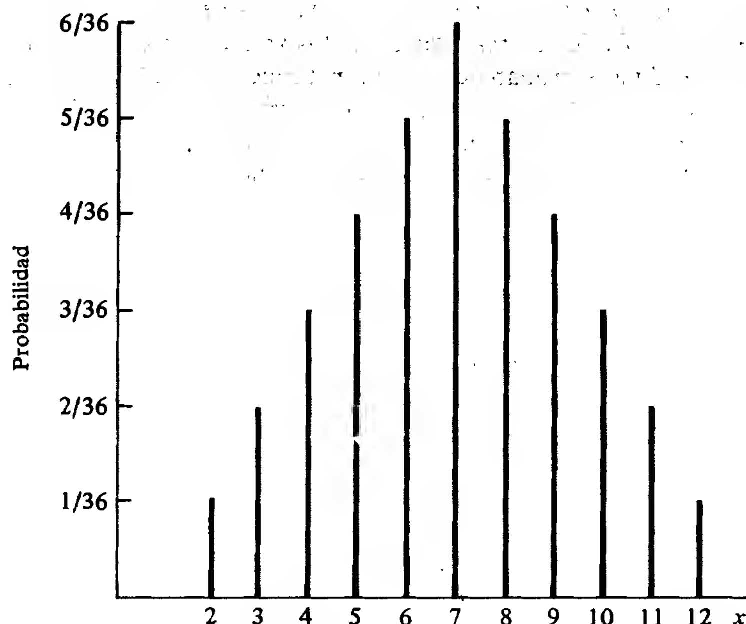


FIGURA 3.1 Probabilidad para las sumas de las caras de dos dados

*ción de probabilidad** de la variable aleatoria X. El término más general, *distribución de probabilidad,* se refiere a la colección de valores de la variable aleatoria y a la distribución de probabilidades entre éstos. Sin embargo, hacer referencia a la distribución de probabilidad de X no sólo implica la existencia de la función de probabilidad, sino también la existencia de la *función de distribución acumulativa* de X.

Definición 3.4 Sea X una variable aleatoria discreta. Se llamará a $p(x) \equiv P(X = x)$ función de probabilidad de la variable aleatoria X, si satisface las siguientes propiedades:

1. $p(x) \geq 0$ para todos los valores x de X;
2. $\Sigma_x \, p(x) = 1$.

Definición 3.5 La función de distribución acumulativa de la variable aleatoria X es la probabilidad de que X *sea menor o igual a un valor específico de* x y está dada por:

$$F(x) \equiv P(X \leq x) = \sum_{x_i \leq x} p(x_i).$$

* El nombre completo de esta función es el de *función másica de probabilidad* de una variable aleatoria discreta.

Por lo tanto, en el caso discreto, una variable aleatoria X está caracterizada por la función de probabilidad puntual $p(x)$, la cual determina la probabilidad puntual de que $X = x$, y por la función de distribución acumulativa $F(x)$, la que representa la suma de las probabilidades puntuales hasta el valor x de X inclusive. Nótese que las definiciones anteriores son consistentes con los axiomas de probabilidad, ya que esta función no es negativa para cualquier valor de la variable aleatoria y la suma de las probabilidades para todos los valores de X es igual a uno.

Ejemplo 3.1 Considérese de nuevo el lanzamiento de dos dados. Si X es la variable aleatoria que representa la suma de las caras, la función de probabilidad de X es

$$p(x) = \begin{cases} \dfrac{6 - |7 - x|}{36} & x = 2, 3, \ldots, 12, \\ 0 & \text{para cualquier otro valor} \end{cases} \tag{3.1}$$

Con (3.1), pueden determinarse las probabilidades para varios valores de X contenidos en la tabla 3.1 y cuya gráfica se muestra en la figura 3.1. Además, puede evaluarse la función de distribución acumulativa de X de la siguiente forma:

$$\begin{aligned}
F(1) &\equiv P(X \leq 1) = 0 \\
F(2) &\equiv P(X \leq 2) = 1/36 \\
F(3) &\equiv P(X \leq 3) = 3/36 \\
F(4) &\equiv P(X \leq 4) = 6/36 \\
F(5) &\equiv P(X \leq 5) = 10/36 \\
F(6) &\equiv P(X \leq 6) = 15/36 \\
F(7) &\equiv P(X \leq 7) = 21/36 \\
F(8) &\equiv P(X \leq 8) = 26/36 \\
F(9) &\equiv P(X \leq 9) = 30/36 \\
F(10) &\equiv P(X \leq 10) = 33/36 \\
F(11) &\equiv P(X \leq 11) = 35/36 \\
F(12) &\equiv P(X \leq 12) = 1.
\end{aligned}$$

Nótese que:

$$\begin{aligned}
&P(X > 7) = 1 - P(X \leq 7) = 1 - F(7) = 15/36; \\
&P(X = 7) = P(X \leq 7) - P(X \leq 6) = F(7) - F(6) = 6/36; \\
&P(5 \leq X \leq 9) = P(X \leq 9) - P(X \leq 4) = F(9) - F(4) = 24/36.
\end{aligned}$$

En general, la función de distribución acumulativa $F(x)$ de una variable aleatoria discreta es una función no decreciente de los valores de X, de tal manera que

1. $0 \leq F(x) \leq 1$ para cualquier x;
2. $F(x_i) \geq F(x_j)$ si $x_i \geq x_j$;
3. $P(X > x) = 1 - F(x)$.

Además, puede establecerse que para variables aleatorias de valor entero se tiene que:

4. $P(X = x) = F(x) - F(x - 1)$;
5. $P(x_i \leq X \leq x_j) = F(x_j) - F(x_i - 1)$.

La gráfica de la distribución acumulativa del ejemplo 3.1 se muestra en la figura 3.2. En esta figura es evidente que la función de distribución acumulativa de una variable aleatoria discreta es una función escalón, que toma un valor superior en cada salto.

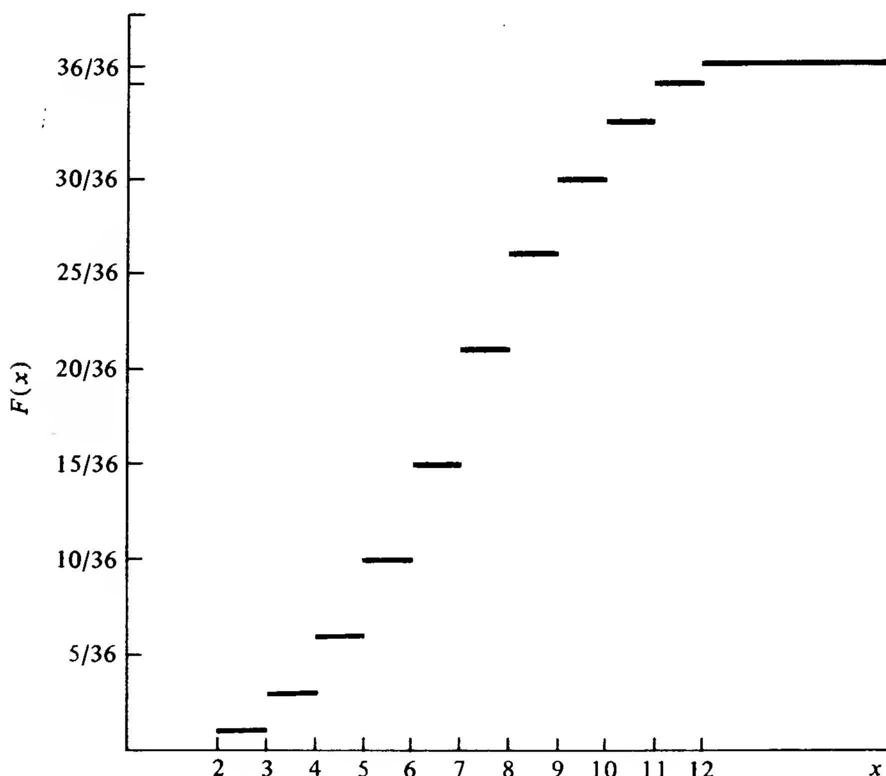


FIGURA 3.2 Representanción gráfica de la función de distribución acumulativa de la suma de las caras de dos dados, cuando éstos se lanzan

3.3 Distribuciones de probabilidad de variables aleatorias continuas

En la sección anterior se trataron distribuciones de probabilidad para variables aleatorias discretas. En ésta se examinarán conceptos similares para variables aleatorias continuas. En el caso discreto, se asignan probabilidades positivas a todos los valores puntuales de la variable aleatoria, pero la suma de todas ellas es uno aún a pesar de que el conjunto de valores sea infinito contable. Para el caso continuo, lo anterior no es posible.

Por esta razón, la probabilidad de que una variable aleatoria continua X tome un valor específico x es cero.

Se ilustrará el sentido de este resultado mediante el siguiente ejemplo: supóngase que se observa el intervalo entre dos llegadas consecutivas a un servicio. Si el dispositivo de medición puede medir el tiempo hasta una décima de segundo, entonces un intervalo de 83.4 seg puede realmente tomarse como la media y el verdadero valor puede encontrarse entre 83.35 y 83.45 seg. Por lo tanto, en el caso continuo es más lógico visualizar las probabilidades de intervalos que de puntos en particular.

La distribución de probabilidad de una variable aleatoria continua X está caracterizada por una función $f(x)$ que recibe el nombre de *función de densidad de probabilidad*. Esta función $f(x)$ no es la misma función de probabilidad que para el caso discreto. Como existe la probabilidad de que X tome el valor específico x es cero, la función de densidad de probabilidad no representa la probabilidad de que $X = x$. Más bien, ésta proporciona un medio para determinar la probabilidad de un intervalo $a \leq X \leq b$.

Para ilustrar lo que se entiende como función de densidad de probabilidad, supóngase que se miden los tiempos, entre dos llegadas consecutivas, de 100 clientes a una tienda y se agrupan en diez intervalos de un minuto cada uno, como se muestra en la tabla 3.2. En este punto se grafican las frecuencias relativas para cada intervalo por medio de rectángulos, como se muestra en la figura 3.3, para indicar que la frecuencia se refiere al intervalo completo más que a un punto en particular del mismo. Nótese que, puesto que la base tiene una longitud igual a uno, el área de cada rectángulo es la frecuencia relativa del correspondiente intervalo y, por lo tanto, la suma de las áreas de todos los rectángulos es igual a uno.

TABLA 3.2 Tiempos entre dos llegadas consecutivas, agrupados, de 100 clientes a un servicio

Intervalo	*Número de llegadas*	*Frecuencia relativa*
$0 < x \leq 1$	22	0.22
$1 < x \leq 2$	18	0.18
$2 < x \leq 3$	17	0.17
$3 < x \leq 4$	13	0.13
$4 < x \leq 5$	14	0.14
$5 < x \leq 6$	8	0.08
$6 < x \leq 7$	6	0.06
$7 < x \leq 8$	7	0.07
$8 < x \leq 9$	3	0.03
$9 < x \leq 10$	2	0.02

Supóngase que en lugar de observar los tiempos entre dos llegadas consecutivas de 100 clientes, se observan los tiempos para 1 000 clientes y se agrupan en 20 intervalos de medio minuto cada uno; o bien pueden observarse los tiempos para 10 000 clientes agrupándolos en 40 intervalos de 15 segundos cada uno. Cada vez que esto se hace, se produce un histograma que es cada vez menos irregular, pero en el que la frecuencia sigue siendo prácticamente la misma. Al continuar este proceso de aumento del número de observaciones mientras se disminuye la amplitud de los intervalos de clase, se llegará a una curva límite. Esto es, cuando el número observado de tiempos, entre dos llegadas consecutivas, sea muy grande y la amplitud de los intervalos de clase sea muy pequeña, la frecuencia relativa aparecerá, en esencia, como una curva lisa. Con base en la figura 3.3, puede especularse que la curva límite para este ejemplo es la que se muestra en la figura 3.4.

La función $f(x)$, cuya gráfica es la curva límite que se obtiene para un número muy grande de observaciones y para una amplitud de intervalo muy pequeña, es la función de densidad de probabilidad para una variable aleatoria continua X, ya que la escala vertical se elige de manera que el área total bajo la curva es igual a uno. La función de densidad de probabilidad de una variable aleatoria continua X se define formalmente de la siguiente manera:

Definición 3.6 Si existe una función $f(x)$ tal que

1. $f(x) \geq 0, \quad -\infty < x < \infty,$

2. $\int_{-\infty}^{\infty} f(x)dx = 1$, y

3. $P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$

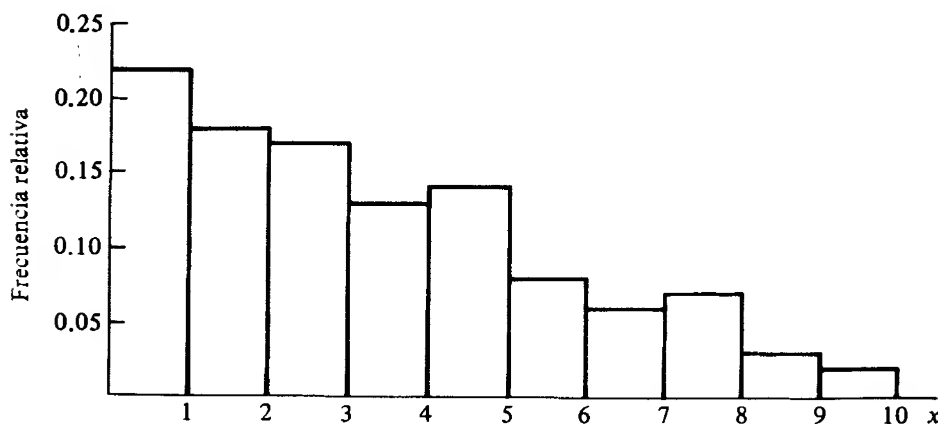


FIGURA 3.3 Frecuencias relativas para los tiempos entre dos llegadas consecutivas, agrupados en diez intervalos

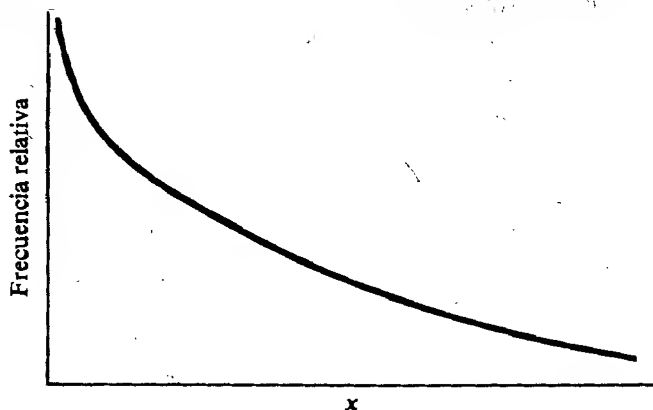


FIGURA 3.4 Curva límite para la frecuencia relativa de los tiempos de llegadas

para cualesquiera a y b, entonces $f(x)$ es la función de densidad de probabilidad de la variable aleatoria continua X.

Puesto que el área total bajo $f(x)$ es uno, la probabilidad del intervalo $a \leq X \leq b$ es el área acotada por la función de densidad y las rectas $X = a$ y $X = b$, como se muestra en la figura 3.5.

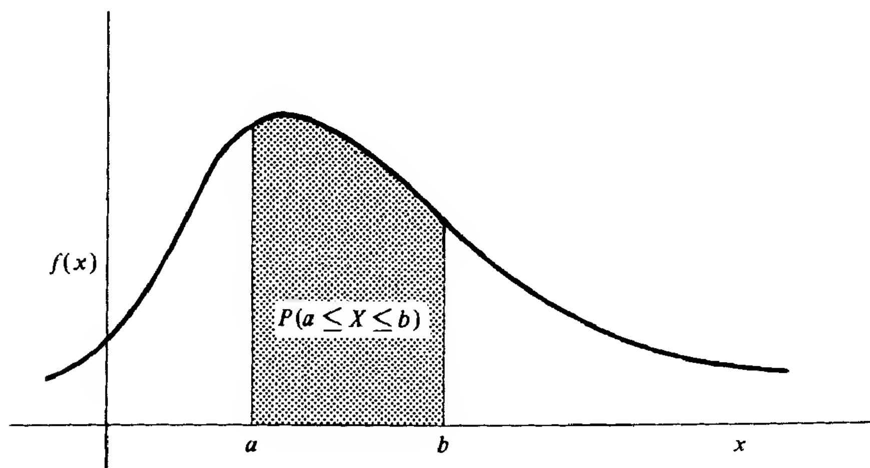


FIGURA 3.5 Probabilidad ilustrada como el área bajo la curva de densidad

Al igual que en el caso de una variable aleatoria discreta, la función de distribución acumulativa de una variable aleatoria continua X es la probabilidad de que X tome un valor menor o igual a algún x específico. Esto es,

$$P(X \leq x) = F(x) = \int_{-\infty}^{x} f(t)dt, \tag{3.2}$$

en donde t es una variable artificial de integración. Por lo tanto, la función de distribución acumulativa $F(x)$ es el área acotada por la función de densidad que se encuentra a la izquierda de la recta $X = x$, como se ilustra en la figura 3.6.

Dado que para cualquier variable aleatoria continua X,

$$P(X = x) = \int_{x}^{x} f(t)dt = 0,$$

entonces:

$$P(X \leq x) = P(X < x) = F(x).$$

La distribución acumulativa $F(x)$, es una función lisa no decreciente de los valores de la variable aleatoria con las siguientes propiedades:

1. $F(-\infty) = 0$;
2. $F(\infty) = 1$;
3. $P(a < X < b) = F(b) - F(a)$;
4. $dF(x)/dx = f(x)$.

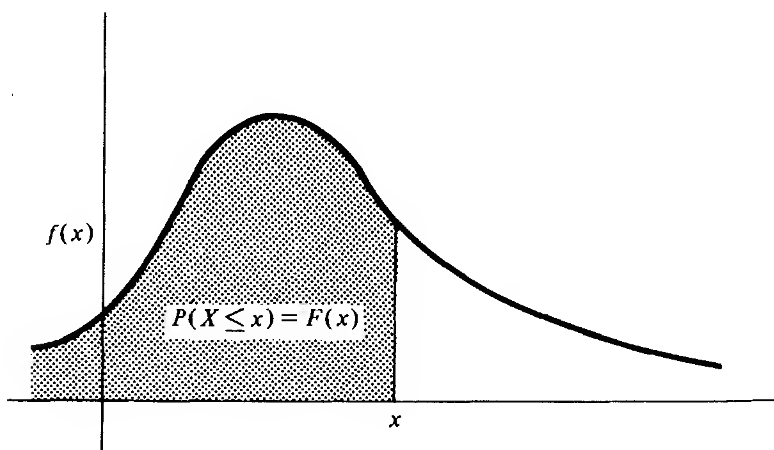


FIGURA 3.6 La distribución acumulativa, ilustrada como un área bajo la curva de densidad

La propiedad de que la derivada de la función de distribución acumulativa es la función de densidad de probabilidad, es una consecuencia del teorema fundamental del cálculo integral.

Ejemplo 3.2 La variable aleatoria X representa el intervalo de tiempo entre dos llegadas consecutivas a una tienda y su función de densidad de probabilidad está dada por:

$$f(x) = \begin{cases} k \exp(-x/2),^{*} & x > 0, \\ 0 & \text{para cualquier otro valor} \end{cases}$$

para una constante k apropiada. Determinar el valor de k, la función de distribución acumulativa, la probabilidad de que $2 < X < 6$, y la probabilidad de que $X \leq 8$. Debe insistirse en que:

$$\int_{-\infty}^{\infty} f(x)dx = 1;$$

por lo tanto, dado que en este ejemplo $f(x) = 0$ si $x \leq 0$, entonces el valor de k está determinado por:

$$k \int_{0}^{\infty} \exp(-x/2)dx = 1.$$

Después de la integración se tiene que:

$$-2k \exp(-x/2)\Big|_{0}^{\infty} = 1,$$

y $k = 1/2$. La función de distribución acumulativa es:

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(t)dt \\ &= \int_{-\infty}^{0} 0dt + \frac{1}{2}\int_{0}^{x} \exp(-t/2)dt \\ &= 1 - \exp(-x/2) \quad \text{para } x > 0, \end{aligned}$$

y $F(x) = 0$ para $x \leq 0$. Además $DF(x)/dx = 1/2 \exp(-x/2)$, que es lo que se esperaba.

La probabilidad de que un intervalo entre dos llegadas consecutivas se encuentre entre dos y seis minutos es:

$$\begin{aligned} P(2 < X < 6) &= \frac{1}{2}\int_{2}^{6} \exp(-x/2)dx = F(6) - F(2) \\ &= [1 - \exp(-3)] - [1 - \exp(-1)] = 0.3181. \end{aligned}$$

* No se dudará en emplear "exp" en lugar de "e", toda vez que esta notación sea menos oscura.

La probabilidad de que transcurran menos de ocho minutos entre dos llegadas consecutivas es:

$$P(X < 8) = F(8) = 1 - \exp(-4) = 0.9817.$$

La probabilidad de que ésta exceda los ocho minutos es $1 - F(8) = \exp(-4) = 0.0183$.

Ejemplo 3.3 La variable aleatoria que representa la proporción de accidentes automovilísticos fatales en Estados Unidos, tiene la siguiente función de densidad:

$$f(x) = \begin{cases} 42x(1-x)^5 & 0 < x \leq 1 \\ 0 & \text{para cualquier otro valor} \end{cases}$$

¿Cuál es la probabilidad de que no más del 25% de los accidentes automovilísticos sean fatales? En otras palabras, ¿cuál es $P[X \leq 0.25]$?

La función $f(x)$ es una densidad de probabilidad dado que:

$$42\int_0^1 x(1-x)^5 dx = 42\left(\frac{x^2}{2} - \frac{5x^3}{3} + \frac{10x^4}{4} - \frac{10x^5}{5} + \frac{5x^6}{6} - \frac{x^7}{7}\right)\Bigg|_0^1 = 1.$$

Nótese que cuando la variable aleatoria X es $1/4$, la función de densidad es $f(1/4) = 2.4917$. De esta forma, en el caso continuo es bastante factible tener, para un valor específico de la variable aleatoria X, un valor de la función de densidad mayor que uno aun a pesar de que la integral de la función de distribución sobre el intervalo completo de valores de la variable aleatoria sea uno. Finalmente, la función de distribución acumulativa es:

$$F(x) = 42\int_0^x t(1-t)^5 dt = 21x^2 - 70x^3 + 105x^4 - 84x^5 + 35x^6 - 6x^7.$$

Por lo tanto, la probabilidad de que la proporción de accidentes automovilísticos fatales sea menor del 25% es:

$$\begin{aligned} F(1/4) &= 21(1/4)^2 - 70(1/4)^3 + 105(1/4)^4 - 84(1/4)^5 + 35(1/4)^6 - 6(1/4)^7 \\ &= 0.5551. \end{aligned}$$

3.4 Valor esperado de una variable aleatoria

El *valor esperado* (o esperanza) de una variable aleatoria es un concepto muy importante en el estudio de las distribuciones de probabilidad. La esperanza de una variable aleatoria tiene sus orígenes en los juegos de azar, debido a que los apostadores deseaban saber cuál era su esperanza de ganar repetidamente un juego. En este sentido, el valor esperado representa la cantidad de dinero promedio que el jugador está dispuesto a ganar o perder después de un número muy grande de apuestas. Este signi-

ficado también es válido para una variable aleatoria. Es decir, el valor promedio de una variable aleatoria después de un número grande de experimentos, es su valor esperado.

Para ilustrar la esencia de la esperanza, se analizará el siguiente juego de azar. Supóngase que se tiene moneda normal y el jugador tiene tres oportunidades para que al lanzarla aparezca una "cara". El juego termina en el momento en el que cae una "cara" o después de tres intentos, lo que suceda primero. Si en el primero, segundo o tercer lanzamiento aparece "cara" el jugador recibe \$2, \$4, y \$8 respectivamente. Si no cae "cara" en ninguno de los tres lanzamientos, pierde \$20. Para determinar la ganancia o pérdida promedio después de un número muy grande de juegos, sea X la variable aleatoria que representa la cantidad que se gana o se pierde cada vez que se juega. Los posibles valores de X junto con sus respectivas probabilidades se encuentran en la tabla 3.3. Después de un número grande de juegos se espera ganar \$2 en cualesquiera de los dos lanzamientos, \$4 en cualesquiera de los cuatro lanzamientos, \$8 una vez cada ocho lanzamientos y se espera perder \$20 una vez en cada ocho intentos. El valor esperado, o la cantidad promedio que se ganaría en cada juego después de un número muy grande de éstos, se determina multiplicando cada cantidad que se gana o se pierde por su respectiva probabilidad y sumando los resultados. De acuerdo con la anterior, la esperanza de ganar es:

$$(\$2)(1/2) + (\$4)(1/4) + (\$8)(1/8) + (-\$20)(1/8) = \$0.50$$

por juego. Nótese que el valor esperado de 50 centavos no es ninguno de los posibles valores de la variable aleatoria; de esta forma, es completamente posible que una variable aleatoria nunca tome el valor de su esperanza.

El ejemplo anterior sugiere la siguiente definición de la esperanza matemática de una variable aleatoria:

Definición 3.7 El valor esperado de una variable aleatoria X es el promedio o valor medio de X y está dado por:

$$E(X) = \sum_x xp(x) \qquad \text{si } x \text{ es discreta, o}$$

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx \quad \text{si } X, \text{ es continua.}$$

en donde $p(x)$ y $f(x)$ son las funciones de probabilidad y de densidad de probabilidad, respectivamente.

TABLA 3.3 Probabilidades de ganar o perder en un juego de azar

X	$P(X)$
2	$P(X = 2) \equiv P(H) = 1/2$
4	$P(X = 4) \equiv P(T \cap H) = 1/4$
8	$P(X = 8) \equiv P(T \cap T \cap H) = 1/8$
−20	$P(X = -20) \equiv P(T \cap T \cap T) = 1/8$

En general, el valor esperado de una función $g(x)$ de la variable aleatoria X, está dado por:

$$E[g(X)] = \sum_x g(x)p(x) \qquad \text{si } x \text{ es discreta, o}$$
$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx \quad \text{si } X, \text{ es continua.} \tag{3.3}$$

La esperanza de una variable aleatoria X no es una función de X sino un número fijo y una propiedad de la distribución de probabilidad de X. Por otra parte, el valor esperado puede no existir dependiendo de si la correspondiente suma o integral no converge en un valor finito.

Ejemplo 3.4 Si la variable aleatoria X representa la suma de las caras de dos dados cuando éstos se lanzan, demostrar que el valor esperado de X es siete.

Con la función de probabilidad de X dada por (3.1) y la definición 3.7, se tiene:

$$E(X) = \sum_{x=2}^{12} xp(x) = (2)(1/36) + (3)(2/36) + \cdots + (12)(1/36) = 7.$$

Ejemplo 3.5 Para el ejemplo 3.3, determinar el valor esperado de la proporción de accidentes fatales en Estados Unidos.
Con la definición 3.7, el valor esperado de la proporción es:

$$\begin{aligned} E(X) &= 42\int_0^1 xf(x)dx \\ &= 42\int_0^1 x^2(1-x)^5dx \\ &= 42x^3\left(\frac{1}{3} - \frac{5x}{4} + 2x^2 - \frac{5x^3}{3} + \frac{5x^4}{7} - \frac{x^5}{8}\right)\Bigg|_0^1 \\ &= 0.25. \end{aligned}$$

Ejemplo 3.6 Supóngase que el tiempo necesario para reparar una pieza de equipo, en un proceso de manufactura, es una variable aleatoria cuya función de densidad de probabilidad es:

$$f(x) = \begin{cases} \frac{1}{5}\exp(-x/5) & x > 0, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

Si la pérdida de dinero es igual al cuadrado del número de horas necesarias para llevar a cabo la reparación, se debe determinar el valor esperado de las pérdidas por reparación.

En este caso es necesario calcular el valor esperado de una función que se encuentra relacionada con la variable aleatoria (el tiempo de reparación). Esta función es:

$$g(x) = x^2;$$

por lo tanto:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx = \frac{1}{5}\int_{0}^{\infty} x^2\exp(-x/5)dx.$$

Para evaluar integrales de este tipo en donde el integrando es un producto de una potencia por una exponencial negativa sobre la recta de los reales positivos, es mejor emplear la función matemática:

$$\Gamma(n) = \int_{0}^{\infty} u^{n-1}\exp(-u)du, \qquad n > 0, \tag{3.4}$$

que se conoce como *función gama* del argumento n. Algunas propiedades de esta función son:

1. $\Gamma(n+1) = n!$ si n es un entero positivo;
2. $\Gamma(n+1) = n\Gamma(n), \qquad n > 0;$
3. $\Gamma(1/2) = \sqrt{\pi}.$

De acuerdo con lo anterior, para evaluar la integral

$$E[g(X)] = \frac{1}{5}\int_{0}^{\infty} x^2\exp(-x/5)dx,$$

en (3.4) es $u = x/5$; en otras palabras, x 5u $dx = 5du$. Entonces:

$$\begin{aligned} E[g(X)] &= \frac{1}{5}\int_{0}^{\infty} x^2\exp(-x/5)dx = \frac{1}{5}\int_{0}^{\infty} (5u)^2\exp(-u)5du \\ &= 25\int_{0}^{\infty} u^2\exp(-u)du \\ &= 25\Gamma(3) \\ &= 50, \end{aligned}$$

50 es el valor esperado de la pérdida por reparación.

Ejemplo 3.7 Un inversionista dispone de $100 000.00 para una inversión de un año. El inversionista está considerando dos opciones: colocar el dinero en el mercado de valores, lo que le garantiza una ganancia anual fija del 15% y un plan de inversión cuya ganancia anual puede considerarse como una variable aleatoria cuyos valores dependen de las condiciones económicas que prevalezcan. Con base en la historia pasada del segundo plan, un analista muy confiable ha determinado los po-

sibles valores de la ganancia y calculado sus probabilidades, como se muestra en la tabla 3.4. Con base en la ganancia esperada ¿cuál de los dos planes debe seleccionarse?

Si se escoge el primer plan, colocar el dinero en el mercado de valores, la ganancia anual que producen $100 mil será de $15 mil, dado que ésta es fija y su valor es del 15%. Para el segundo plan, sea X la variable aleatoria que representa la ganancia. Con la definición 3.7, se tiene:

$$E(X) = (0.3)(0.2) + (0.25)(0.2) + \cdots + (0.05)(0.05) = 0.205.$$

De acuerdo con lo anterior, el segundo plan es una elección mucho mejor puesto que ofrece una ganancia esperada de $20 500. Sin embargo, el lector debe ser cauteloso en este punto, dado que el valor de $20 500 es únicamente un valor esperado y el inversionista no tiene ninguna garantía dc que su ganancia real se encuentre cercana a este valor.

A continuación se enunciarán y demostrarán algunas propiedades importantes de la esperanza de una variable aleatoria. Se usará el caso continuo, a pesar de que estas propiedades también son válidas para variables aleatorias discretas. Sea X una variable aleatoria continua con una función de densidad de probabilidad $f(x)$.

1. El valor esperado de una constante c es el valor de la constante.

$$E(c) = \int_{-\infty}^{\infty} cf(x)dx = c\int_{-\infty}^{\infty} f(x)dx = c.$$

2. El valor esperado de la cantidad $aX + b$, en donde a y b son constantes, es el producto de a por el valor esperado de x más b.

$$\begin{aligned} E(aX + b) &= \int_{-\infty}^{\infty} (ax + b)f(x)dx = a\int_{-\infty}^{\infty} xf(x)dx + b\int_{-\infty}^{\infty} f(x)dx \\ &= aE(X) + b. \end{aligned}$$

3. El valor esperado de la suma de dos funciones $g(X)$ y $h(X)$ de X es la suma de los valores esperados de $g(X)$ y $h(X)$.

$$E[g(X) + h(X)] = \int_{-\infty}^{\infty} [g(x) + h(x)]f(x)dx$$

TABLA 3.4 Valores de la ganancia para el ejemplo 3.7

Ganancia (%)	*Probabilidad*
30	0.20
25	0.20
20	0.30
15	0.15
10	0.10
5	0.05

$$= \int_{-\infty}^{\infty} g(x)f(x)dx + \int_{-\infty}^{\infty} h(x)f(x)dx$$

$$= E[g(X)] + E[h(X)].$$

3.5 Momentos de una variable aleatoria

Los *momentos* de una variable aleatoria X* son los valores esperados de ciertas funciones de X. Éstos forman una colección de medidas descriptivas que pueden emplearse para caracterizar la distribución de probabilidad de X y especificarla si todos los momentos de X son conocidos. A pesar de que los momentos de X pueden definirse alrededor de cualquier punto de referencia, generalmente se definen alrededor del cero o del valor esperado de X. El uso de los momentos de una variable aleatoria para caracterizar a la distribución de probabilidad es una tarea muy útil. Lo anterior es especialmente cierto en un medio en el que es poco probable que el experimentador conozca la distribución de probabilidad. Todas las proposiciones con respecto a los momentos se encuentran sujetas a la existencia de las sumas o integrales que las definan.

Definición 3.8 Sea X una variable aleatoria. El r-ésimo momento de X alrededor del cero se define por:

$$\mu_r' = E(X^r) = \sum_x x^r p(x) \qquad \text{si } X \text{ es discreta, o}$$

$$\mu_r' = E(X^r) = \int_{-\infty}^{\infty} x^r f(x)dx \qquad \text{si } X \text{ es continua.}$$

El primer momento alrededor del cero es la *media* o *valor esperado* de la variable aleatoria y se denota por μ; de esta manera se tiene que $\mu_1' = \mu = E(X)$. Con base en el material del capítulo uno, la media de una variable aleatoria se considera como una cantidad numérica alrededor de la cual los valores de la variable aleatoria tienden a agruparse. Por lo tanto, la media es una medida de tendencia central.

Definición 3.9 Sea X una variable aleatoria. El r-ésimo momento central de X o el r-ésimo momento alrededor de la media de X se define por:

$$\mu_r = E(X - \mu)^r = \sum_x (x - \mu)^r p(x) \qquad \text{si } X \text{ es discreta, o}$$

$$\mu_r = E(X - \mu)^r = \int_{-\infty}^{\infty} (x - \mu)^r f(x)dx \qquad \text{si } X \text{ es continua.}$$

* También es apropiado emplear la frase *momentos de la distribución de probabilidad de X*.

El momento central cero de cualquier variable aleatoria es uno, dado que:

$$\mu_0 = E(X - \mu)^0 = E(1) = 1.$$

De manera similar, el primer momento central de cualquier variable aleatoria es cero, dado que:

$$\mu_1 = E(X - \mu) = E(X) - \mu = 0.$$

El segundo momento central:

$$\mu_2 = E(X - \mu)^2,$$

recibe el nombre de *varianza* de la variable aleatoria. Puesto que:

$$\begin{aligned} \mu_2 = Var(X) &= E(X - \mu)^2 \\ &= E(X^2 - 2X\mu + \mu^2) \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= \mu_2' - \mu^2, \end{aligned} \tag{3.5}$$

la varianza de cualquier variable aleatoria es el segundo momento alrededor del origen menos el cuadrado de la media. Generalmente se denota por σ^2. La varianza de una variable aleatoria es una medida de la dispersión de la distribución de probabilidad de ésta. Por ejemplo, en el caso continuo si la mayor parte del área por debajo de la curva de distribución se encuentra cercana a la media, la varianza es pequeña; si la mayor parte del área se encuentra muy dispersa alrededor de la media, la varianza será grande. La raíz cuadrada positiva de la varianza recibe el nombre de ***desviación estándar*** y se denota por σ. A pesar de que σ^2 y σ son los símbolos más universales para la varianza y la desviación estándar, respectivamente; en este libro no se dudará en emplear las notaciones $\sigma^2(X)$ o $Var(X)$ para la varianza y $\sigma(X)$ o $d.e.\ (X)$ para la desviación estándar dada su identificación explícita con la variable aleatoria involucrada. Por la misma razón, a veces será necesario emplear la notación $\mu_r(X)$ para denotar el r-ésimo momento central de X.

Es útil notar que la varianza de una variable aleatoria X es invariable; es decir, $Var(X + b) = Var(X)$ para cualquier constante b. De manera más general, se demostrará que $Var(aX + b) = a^2Var(X)$ para cualesquiera dos contantes a y b. Por definición,

$$\begin{aligned} Var(aX + b) &= E(aX + b)^2 - E^2(aX + b) \\ &= E(a^2X^2 + 2abX + b^2) - [aE(X) + b]^2 \\ &= a^2E(X^2) + 2abE(X) + b^2 - a^2E^2(X) - 2abE(X) - b^2 \\ &= a^2E(X^2) - a^2E^2(X) \\ &= a^2[E(X^2) - E^2(X)] \\ &= a^2\, Var\,(X). \end{aligned}$$

Una medida que compara la dispersión relativa de dos distribuciones de probabilidad es el *coeficiente de variación,* que está definido por:

$$V = \sigma/\mu. \tag{3.6}$$

El coeficiente de variación expresa la magnitud de la dispersión de una variable aleatoria con respecto a su valor esperado. V es una medida estandarizada de la variación con respecto a la media, especialmente útil para comparar dos distribuciones de probabilidad cuando la escala de medición difiere de manera apreciable entre éstas. Por ejemplo, dadas las variables aleatorias X y Y, supóngase que:

$$E(X) = 120, \ Var(X) = 36; E(Y) = 40, \ Var(Y) = 16.$$

A pesar de que la dispersión de X, por su desviación estándar, es más grande que la de Y, en un sentido absoluto, la dispersión relativa de X es menor que la dispersión relativa de Y, puesto que:

$$V_X = 6/120 = 0.05,$$

pero:

$$V_Y = 4/40 = 0.10.$$

Por lo tanto, la distribución de probabilidad de Y muestra una mayor dispersión relativa con respecto a la media que la distribución correspondiente a X.

En este punto, se examinarán los momentos centrales tercero y cuarto de una variable aleatoria X. Estos momentos centrales proporcionan información muy útil con respecto a la forma de la distribución de probabilidad de X. A pesar de que pueden considerarse momentos de orden superior, su utilidad para caracterizar una distribución de probabilidad es mucho menor que la de los primeros cuatro momentos. El tercer momento central

$$\mu_3 = E(X - \mu)^3, \tag{3.7}$$

está relacionado con la *asimetría* de la distribución de probabilidad de X. Ya se demostró que el segundo momento central (la varianza) puede expresarse en términos de los primeros dos momentos alrededor del cero. De hecho, cualquier momento central de una variable aleatoria X puede expresarse en términos de los momentos de ésta, alrededor del cero. Por definición:

$$\mu_r = E(X - \mu)^r,$$

pero la expansión de $(X - \mu)^r$ puede expresarse como:

$$(X - \mu)^r = \sum_{i=0}^{r} (-1)^i \frac{r!}{(r - i)!i!} \mu^i x^{r-i}.$$

Ya que la esperanza de una suma es igual a la suma de las esperanzas, se tiene que:

$$\mu_r = \sum_{i=0}^{r} (-1)^i \frac{r!}{(r-i)!i!} \mu^i E(X^{r-i})$$

$$= \sum_{i=0}^{r} (-1)^i \frac{r!}{(r-i)!i!} \mu^i \mu'_{r-i}.$$

En particular,

$$\mu_3 = \mu'_3 - 3\mu\mu'_2 + 2\mu^3. \tag{3.8}$$

Para las distribuciones de probabilidad que presentan un solo pico, si $\mu_3 < 0$, se dice que la distribución es *asimétrica negativamente;* si $\mu_3 > 0$, la distribución es *asimétrica positivamente;* y si $\mu_3 = 0$, la distribución recibe el nombre de *simétrica.* Sin embargo, a menos que la distribución presente un solo pico, el conocimiento de μ_3 no es suficiente para tener una idea de la forma de la distribución. Aun así, el tercer momento central puede dar resultados erróneos, dado que depende de las unidades en las que se mide la variable aleatoria X. Para estos casos, una medida más apropiada de la asimetría, es el tercer momento estandarizado, dado por;

$$\alpha_3^* = \mu_3/(\mu_2)^{3/2}, \tag{3.9}$$

que recibe el nombre de *coeficiente de asimetría.* El coeficiente α_3 es la medida de la asimetría de una distribución de probabilidad con respecto a su dispersión. Una dis-

* En ocasiones, será necesario identificar a la variable aleatoria explícitamente, con el propósito de evitar ambigüedades.

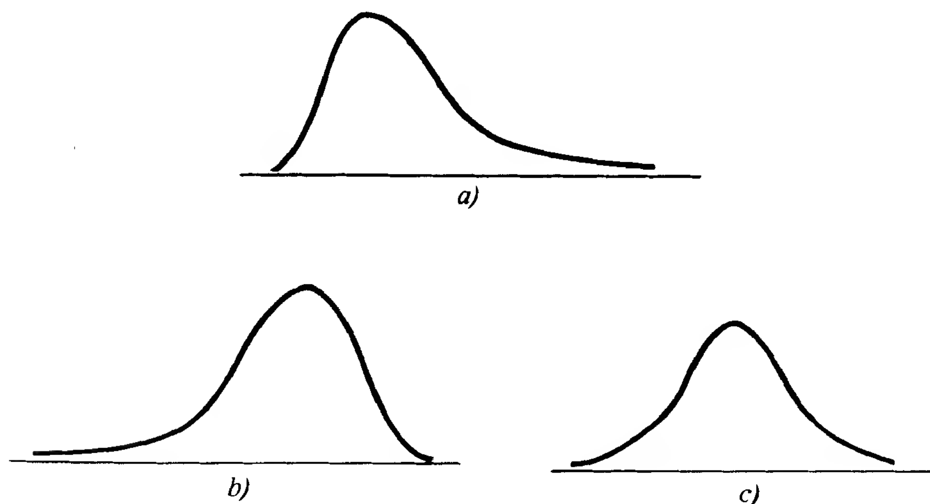


FIGURA 3.7 Funciones de densidad de probabilidad típicas de distribuciones: *a)* asimétrica positivamente, *b)* asimétrica negativamente y *c)* simétrica.

tribución de probabilidad es asimétrica positiva, negativa o simétrica si $\alpha_3 > 0$, $\alpha_3 < 0$, o $\alpha_3 = 0$ respectivamente, como se muestra en la figura 3.7. Nótese que si la distribución de probabilidad de una variable aleatoria X es simétrica, todos los momentos centrales de X de orden impar serán cero, dado que cada valor positivo de $(X - \mu)^r$ se cancela por un valor negativo de la misma magnitud y de igual probabilidad.

El cuarto momento central,

$$\begin{aligned} \mu_4 &= E(X - \mu)^4 \\ &= \mu_4' - 4\,\mu\,\mu_3' + 6\,\mu^2\mu_2' - 3\,\mu^4, \end{aligned} \tag{3.10}$$

es una medida de qué tan *puntiaguda* es la distribución de probabilidad y recibe el nombre de *curtosis*. Al igual que para el tercer momento, es preferible emplear el cuarto momento estandarizado,

$$\alpha_4 = \mu_4/\mu_2^2, \tag{3.11}$$

como una medida relativa de la curtosis. Si $\alpha_4 > 3$, la distribución de probabilidad presenta un pico relativamente alto y recibe el nombre de *leptocúrtica;* si $\alpha_4 < 3$, la distribución es relativamente plana y recibe el nombre de *platicúrtica;* y si $\alpha_4 = 3$, la distribución no presenta un pico muy alto ni muy bajo y recibe el nombre de *mesocúrtica*. Los tres tipos de distribuciones se encuentran ilustrados en la figura 3.8.

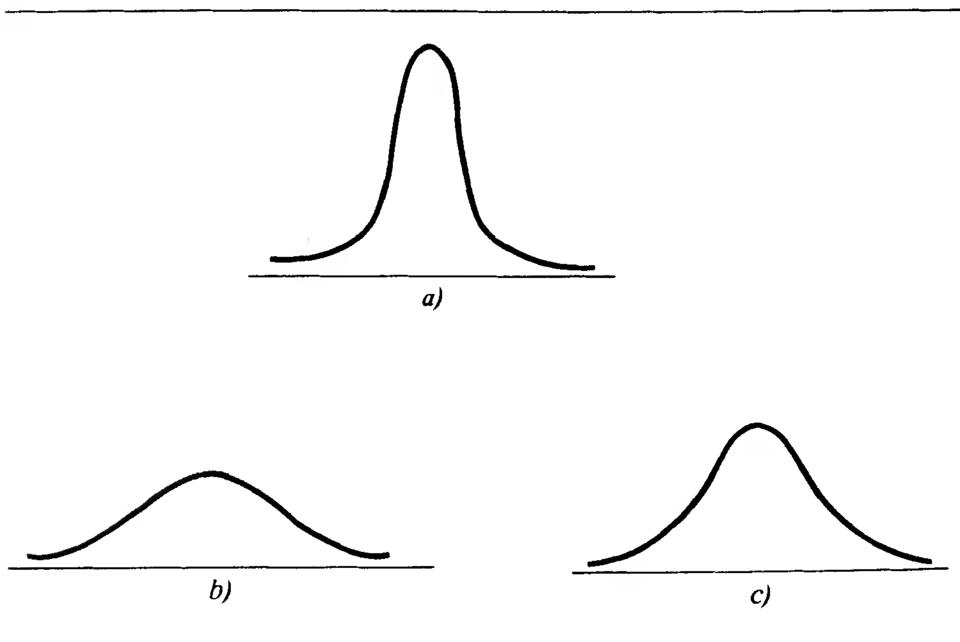


FIGURA 3.8 Funciones de densidad de probabilidad típicas de distribuciones: *a)* leptocúrticas, *b)* platicúrticas y *c)* mesocúrticas

El valor de tres se emplea como una referencia debido a que en la práctica la curtosis estandarizada de una distribución de probabilidad se compara con la de una distribución ampliamente utilizada, conocida como distribución normal, cuyo valor es tres. La distribución normal se estudia con gran detalle prosteriormente.

Los momentos estandarizados tercero y cuarto, también se conocen como los factores de forma primero y segundo, respectivamente, de la distribución de probabilidad debido a que, en gran medida, determinan la forma de la distribución de probabilidad.

Ejemplo 3.8 Dos vendedores de seguros de vida, A y B, visitan de ocho a 12 clientes potenciales por semana, respectivamente. Sean X y Y dos variables aleatorias que representan el número de sendos seguros vendidos por A y B, como resultado de las visitas. Con base en una gran cantidad de información pasada, las probabilidades para los valores de X y Y son las siguientes:

x	0	1	2	3	4	5	6	7	8
$p(x)$	0.02	0.09	0.21	0.28	0.23	0.12	0.04	0.01	0

y	0	1	2	3	4	5	6	7	8	9	10	11	12
$p(y)$	0.06	0.21	0.28	0.24	0.13	0.05	0.02	0.01	0	0	0	0	0

Comparar y contrastar las distribuciones de probabilidad de X y Y empleando sus medias, varianzas y factores de forma.

Con base en la definición 3.8, los primeros cuatro momentos de X alrededor del cero son:

$$\mu = (0)\,(0.2) + (1)\,(0.09) + \cdots + (8)(0) \quad = 3.18$$

$$\mu_2' = (0)^2(0.02) + (1)^2(0.09) + \cdots + (8)^2(0) = 12.06$$

$$\mu_3' = (0)^3(0.02) + (1)^3(0.09) + \cdots + (8)^3(0) = 51.12$$

$$\mu_4' = (0)^4(0.02) + (1)^4(0.09) + \cdots + (8)^4(0) = 235.86.$$

Al emplear las expresiones 3.5, 3.8 y 3.10, respectivamente, se determina que $Var(X) = 1.95$, $\mu_3(X) = 0.3825$ y $\mu_4(X) = 10.565$. Los primeros dos factores de forma de la distribución de probabilidad de X se obtienen empleando (3.9) y (3.11), respectivamente, y son $\alpha_3(X) = 0.1405$ y $\alpha_4(X) = 2.78$.

Con el mismo procedimiento, los primeros cuatro momentos de Y alrededor del cero son $\mu = 2.45$, $\mu_2' = 8.03$, $\mu_3' = 31.25$ y $\mu_4' = 138.59$. De esta manera $Var(Y) = 2.03$, $\mu_3(Y) = 1.6418$, $\mu_4(Y) = 13.4504$, $\alpha_3(Y) = 0.5676$, y $\alpha_4(Y) = 3.26$.

A primera vista, parece existir muy poca diferencia entre las distribuciones de X y Y con respecto a la media y la varianza, pero la distribución de Y tiene un sesgo positi-

vo más grande que la de X. Además, la distribución de X es platicúrtica ($\alpha_4 < 3$), mientras que la de Y es leptocúrtica ($\alpha_4 > 3$).

En este momento se considerará el concepto de variable aleatoria estandarizada. Sea X cualquier variable aleatoria con media μ y desviación estándar σ. La cantidad

$$Y = (X - \mu)/\sigma \tag{3.12}$$

define una variable aleatoria Y con media cero y desviación estándar uno. Esta variable aleatoria recibe el nombre de *variable estandarizada* correspondiente a X. De hecho, para cualquier valor particular x de X el valor $y = (x - \mu)/\sigma$ indica la desviación del valor x del valor esperado de X en términos de las unidades de la desviación estándar. Por ejemplo, si X representa la calificación de una prueba de inteligencia, y si $E(X) = 100$ y $Var(X) = 100$, entonces $Y = (X - 100)/10$ es la variable estandarizada correspondiente a X. Además, si una persona posee un coeficiente intelectual de 120, entonces se encontrará a dos desviaciones estándar del coeficiente intelectual medio.

El valor esperado de Y es cero, puesto que:

$$E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma} E(X - \mu) = 0.$$

De hecho, puesto que $E(Y) = 0$, el r-ésimo momento central de Y es:

$$\begin{aligned} \mu_r(Y) = E(Y^r) &= E\left(\frac{X - \mu}{\sigma}\right)^r \\ &= \frac{1}{\sigma^r} E(X - \mu)^r \\ &= \mu_r(X)/\sigma^r; \end{aligned}$$

de esta manera se tiene que:

$$\mu_r(Y) = \mu_r(X)/\{\mu_2(X)\}^{r/2}. \tag{3.13}$$

De (3.13) es evidente que $Var(Y) = \mu_2(Y) = 1$. En particular, nótese que $\alpha_3(Y) = \alpha_3(X)$ y $\alpha_4(Y) = \alpha_4(X)$. La estandarización de una variable aleatoria afecta a la media y a la varianza, pero no a los factores de forma.

Ejemplo 3.9 Considérense las variables aleatorias X y Y, cuyas funciones de densidad de probabilidad son

$$f(x) = \begin{cases} 1/30 & 80 \leq x \leq 110, \\ 0 & \text{para cualquier otro valor} \end{cases}$$

$$\text{y} \quad f(y) = \begin{cases} \dfrac{1}{10\,000} \exp(-y/10\,000) & y > 0, \\ 0 & \text{para cualquier otro valor;} \end{cases}$$

Determinar y comparar la media, la varianza y los momentos estandarizados tercero y cuarto, de X y Y.

El principal objetivo de este problema es contrastar las distribuciones de probabilidad de X y de Y, mediante la comparación de sus cuatro primeros momentos y, en alguna medida, proporcionar un análogo teórico de los ejemplos 1.1 y 1.2. El lector puede verificar, de manera fácil, que las distribuciones de probabilidad de X y Y son muy diferentes, graficando las correspondientes funciones de densidad. Como se verá, gran parte de la diferencia puede descubrirse a través de las comparaciones entre los cuatro primeros momentos de X y Y.

Para facilitar los cálculos, sea $c_1 = 1/30$ y $c_2 = 1/10\,000$. Para la variable aleatoria X:

$$E(X) = c_1 \int_{80}^{110} x dx = \frac{c_1}{2} x^2 \Big|_{80}^{110} = 95$$

y

$$Var(X) = c_1 \int_{80}^{110} (x - 95)^2 dx = c_1 \int_{-15}^{15} u^2 du = 75,$$

en donde $u = x - 95$ y $dx = du$. Por lo tanto, se tiene que $d.e.(X) = 8.66$. Para los momentos de orden superior:

$$E(X - 95)^3 = c_1 \int_{80}^{110} (x - 95)^3 dx = c_1 \int_{-15}^{15} u^3 du = 0$$

y

$$E(X - 95)^4 = c_1 \int_{80}^{110} (x - 95)^4 dx = c_1 \int_{-15}^{15} u^4 du = 10\,125.$$

De acuerdo con (3.9) y (3.11), los factores de forma, primero y segundo, de X son $\alpha_3(X) = 0/(75)^{3/2} = 0$ y $\alpha_4(X) = 10{,}125/5{,}625 = 1.8$, respectivamente. La distribución de probabilidad de X es simétrica y está centrada alrededor del valor 95, tiene una varianza de 75 y una desviación estándar de 8.66, y tiende a ser plana en su parte superior.

Para la variable aleatoria Y:

$$E(Y) = c_2 \int_0^{\infty} y \exp(-c_2 y) dy = c_2 \int_0^{\infty} \frac{1}{c_2} u \exp(-u) \frac{1}{c_2} du = \Gamma(2)/c_2 = 10\,000$$

y

$$E(Y^2) = c_2 \int_0^{\infty} y^2 \exp(-c_2 y) dy = \Gamma(3)/c_2^2 = 2 \times 10^8,$$

en donde $u = c_2 y$ y $dy = du/c_2$. De esta manera se tiene que $Var(Y) = 1 \times 10^8$, y $d.e.(X) = 10\,000$. Además:

$$E(Y^3) = c_2 \int_0^{\infty} y^3 \exp(-c_2 y) dy = \Gamma(4)/c_2^3 = 6 \times 10^{12}.$$

Con (3.8) y (3.9) se determina que $E(Y - 10\,000)^3 = 2 \times 10^{12}$, y $d.e.(Y) = 10\,000$. De manera similar:

$$E(Y^4) = c_2 \int_0^\infty y^4 \exp(-c_2 y)dy = \Gamma(5)/c_2^4 = 24 \times 10^{16}.$$

Con (3.10) y (3.11), respectivamente, se obtiene que $E(Y - 10\,000)^4 = 9 \times 10^{16}$, y $\alpha_4(Y) = 9$. Puede concluirse que la distribución de Y está sesgada positivamente, tiene un pico relativamente alto, una media de 10 000, una varianza de 1×10^8, y una desviación estándar de 10 000.

3.6 Otras medidas de tendencia central y dispersión

A pesar de que la media y la varianza son las principales medidas de tendencia central y dispersión, existen otras medidas empleadas comúnmente. Se debe recordar que en el capítulo uno, la mediana y la moda eran otras medidas útiles de tendencia central.

Definición 3.10 Para cualquier variable aleatoria X, se define a la mediana $x_{0.5}$ de X, para ser:

$$P(X < x_{0.5}) \leq 1/2 \quad \text{y} \quad P(X \leq x_{0.5}) \geq 1/2 \qquad \text{si } X \text{ es discreta, o}$$

$$P(X \leq x_{0.5}) = 1/2 \qquad \text{si } X \text{ es continua.}$$

Si existe uno de estos valores para X, entonces $x_{0.5}$ recibe el nombre de mediana de la distribución de X. La mediana es una medida de tendencia central, en el sentido de que es el valor para el cual la distribución de probabilidad se divide en dos partes iguales.

Definición 3.11 Para cualquier variable aleatoria X, se define la moda como el valor x_m de X que maximiza la función de probabilidad, si X es discreta, o la función de densidad si X es continua.

Si existe uno de estos valores para X, entonces x_m recibe el nombre de moda de la distribución de X. Si X es continua la moda es la solución de $df(x)/dx = 0$ si $d^2f(x)/dx^2 < 0$. Si la segunda derivada es positiva, el valor recibe el nombre de antimoda; éste se encuentra en las distribuciones que tienen forma de U. Si existen varios máximos o mínimos, las distribuciones de probabilidad reciben el nombre de multimodales.

De acuerdo con la exposición empírica del capítulo uno, la media de una variable aleatoria es generalmente la medida preferida de tendencia central. Sin embargo, en algunas situaciones la mediana, y en menor grado la moda, pueden ser medidas de tendencia central mucho más apropiadas. Por ejemplo, en distribuciones unimodales cuya asimetría es grande, el valor esperado de la variable aleatoria puede verse afectado por los valores extremos de la distribución, mientras que la mediana no lo

estará. Para distribuciones unimodales con asimetría negativa, la mediana es más grande que la media, mientras que lo opuesto es cierto para distribuciones unimodales con asimetría positiva. Para distribuciones unimodales simétricas, la media, mediana y moda coinciden en valor.

Ejemplo 3.10 Sea X una variable alaeatoria que representa el tiempo de duración, en horas, de un cierto componente eléctrico. Si la función de densidad de probabilidad de X está dada por

$$f(x) = \begin{cases} \dfrac{1}{1000}\exp(-x/1000) & x > 0, \\ 0 & \text{para cualquier otro valor,} \end{cases}$$

determinar y comparar la media y la mediana.

La media de X es:

$$E(X) = \frac{1}{1000}\int_0^{\infty} x\exp(-x/1000)dx = 1000\int_0^{\infty} u\exp(-u)du$$
$$= 1000\Gamma(2) = 1000 \text{ horas.}$$

en donde $x = 1\,000u$ y $dx = 1\,000du$. La mediana de X es:

$$P(X \leq x_{0.5}) \equiv F(x_{0.5}) = \frac{1}{1000}\int_0^{x_{0.5}} \exp(-x/1000)dx = 0.5$$
$$= 1 - \exp(-x_{0.5}/1000) = 0.5.$$

Por lo tanto,

$$x_{0.5} = -1000\ln(0.5) = 693.15 \text{ horas.}$$

Se puede demostrar que esta función de probabilidad es asimétrica positivamente, puesto que su coeficiente de asimetría es $\alpha_3 = 2$. De esta forma, la duración media de 1 000 horas se encuentra afectada por los valores de la variable aleatoria en los extremos de la distribución. De hecho la probabilidad de que un componente trabaje más que el valor promedio, es de 0.3679 puesto que

$$P(X > \mu) = 1 - F(\mu) = 1 - 0.6321 = 0.3679.$$

En este caso, el valor de la mediana para el tiempo de duración, 693.15 hr, resulta ser una medida más apropiada de tendencia central.

Además de la varianza, existen otras medidas de dispersión para variables aleatorias como el recorrido interdecil, el recorrido intercuartil y la desviación media, como se mencionó en el capítulo uno. Los primeros dos son funciones de los cuantiles de la distribución de probabilidad. La desviación media es el paralelo conceptual de la desviación estándar, con excepción de que se emplea el valor absoluto de la diferencia entre el valor de la variable aleatoria y su valor esperado en lugar del cuadrado de ésta.

Definición 3.12 Para cualquier variable aleatoria X, el valor cuantil x_q de orden q, $0 < q < 1$, es el valor de X tal que:

$$P(X < x_q) \leq q \quad \text{y} \quad P(X \leq x_q) \geq q \quad \text{si } X \text{ es discreta, o}$$

$$P(X \leq x_q) = q \quad \text{si } X \text{ es continua.}$$

Generalmente los valores cuantiles de una variable aleatoria continua son relativamente fáciles de determinar. Sin embargo, para variables aleatorias discretas los valores cuantiles generalmente se obtienen por interpolación, dado que no siempre es posible obtener una solución exacta.

Los cuantiles utilizados comúnmente son los percentiles, deciles y cuartiles. Los percentiles son los puntos que dividen a la distribución de probabilidad en 100 intervalos, cada uno con probabilidad 0.01; los deciles y cuartiles son los puntos que dividen a la distribución de probabilidad en 10 y cuatro intervalos, cada uno con probabilidad de 0.1 y 0.25, respectivamente. Nótese que la mediana es también el cincuentavo porcentil, el quinto decil y el segundo cuartil.

El recorrido interdecil es la diferencia entre el noveno y primer decil, y el recorrido intercuartil es la diferencia entre el tercer y primer cuartil. De esta manera el recorrido interdecil es una medida de la dispersión de la mitad del 80% de la distribución de probabilidad, en tanto que el recorrido intercuantil refleja la variación de la mitad del 50% de la distribución. En ambos casos, al excluir los efectos de los valores extremos de la distribución, se tiene la capacidad de medir la variabilidad de una variable aleatoria alrededor de la mitad de su distribución de probabilidad.

Los recorridos interdecil e intercuartil, son dos medidas de dispersión que se emplean en disciplinas como educación, economía, finanzas e ingeniería. El recorrido interdecil se emplea muchas veces en pruebas educacionales para medir la variabilidad en el desempeño sin importar los valores por arriba o por debajo de un 10% de un valor predeterminado. El recorrido intercuartil se emplea en muchas ocasiones, en economía y finanzas, para medir la variabilidad de una variable aleatoria alrededor de una porción de su distribución de probabilidad.

Definición 3.13 La desviación media de una variable aleatoria X es el valor esperado de la diferencia absoluta entre X y su media, y está dado por:

$$E|X - \mu| = \sum_{\text{toda } x} |x - \mu| p(x) \quad \text{si } X \text{ es discreta, o}$$

$$E|X - \mu| = \int_{-\infty}^{\infty} |x - \mu| f(x) dx \quad \text{si } X \text{ es continua,}$$

A pesar de que la desviación media es una medida legítima de dispersión, existen distribuciones de probabilidad para las que dar un tratamiento analítico es o muy difícil o imposible. A pesar de todo y como se ilustró en el capítulo uno, la desviación media es una alternativa viable a la desviación estándar como medida de dispersión para conjuntos de datos cuyo fundamento se encuentra en evidencia empírica. Debe notarse que para distribuciones con valores grandes en sus extremos, el valor de la

desviación media se ve menos afectado que la desviación estándar por la existencia de valores extremos.

Ejemplo 3.11 Supóngase que en cierto proceso de llenado, la desviación entre el peso verdadero de un recipiente con respecto al valor específico, es una variable aleatoria Z, cuya función de densidad de probabilidad está dada por

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2) \qquad -\infty < z < \infty.$$

Determinar la media, la desviación estándar, el recorrido interdecil, el recorrido intercuartil y la desviación media de Z.

Como se verá en el capítulo cinco, esta función de densidad es un miembro especial de una familia muy útil en las distribuciones que reciben el nombre de familia normal o Gausiana. De hecho, la función de distribución acumulativa de Z se encuentra bien tabulada, como puede observarse en la tabla D del apéndice. Además, como se verá posteriormente:

$$E(Z) = 0,\ Var(Z) = 1, \text{ y } d.e.(Z) = 1.$$

Para determinar el recorrido interdecil, los valores cuantiles $z_{0.1}$ y $z_{0.9}$ se encuentran definidos por:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_{0.1}} \exp(-t^2/2)dt = 0.1 \quad \text{y} \quad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_{0.9}} \exp(-t^2/2)dt = 0.9$$

y se obtienen de la tabla D*; sus valores son $z_{0.1} = -1.28$ y $z_{0.9} = 1.28$. El recorrido interdecil es $z_{0.9} - z_{0.1} = 2.56$ En otras palabras, el 80% de todos los recipientes presentarán una desviación no mayor de 1.28 unidades, en cualquier dirección del peso especificado. De manera similar, a partir de la tabla D los valores cuantiles $z_{0.25}$ y $z_{0.75}$ son –0.675 y 0.675 respectivamente. Por lo tanto, el recorrido intercuartil es $z_{0.25} = -0.675$ $z_{0.75} - z_{0.25} = 1.35$ unidades.

Puesto que para la desviación mediana $E(Z) = 0$, se tiene:

$$\begin{aligned} E|Z| &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |z| \exp(-z^2/2)dz \\ &= \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} z \exp(-z^2/2)dz \\ &= -\frac{2}{\sqrt{2\pi}} \exp(-z^2/2) \Big|_0^{\infty} \\ &= 2/\sqrt{2\pi} \\ &= 0.7979 \text{ unidades.} \end{aligned}$$

* El uso de la tabla *D* se explica con mucho detalle en el capitulo cinco.

Nótese que dado que la desviación estándar es uno, el recorrido interdecil es de aproximadamente 2.56 unidades de la desviación estándar, el recorrido intercuartil es de 1.35 unidades de la desviación estándar y la desviación media tiene un valor de aproximadamente 0.7979 unidades de la desviación estándar. Los resultados anteriores son siempre válidos para la familia de distribuciones normales.

El siguiente ejemplo ilustra una situación teórica, en la que se tiene una distribución con algunos valores muy grandes y para la cuál la mediana, el recorrido interdecil y el recorrido intercuartil son medidas de tendencia central y dispersión más apropiadas que la media y la varianza.

Ejemplo 3.12 Sea X una variable aleatoria cuya función de densidad de probabilidad está dada por:

$$f(x) = \begin{cases} \frac{1}{8}x^{-1/2}\exp(-x^{1/2}/4) & x > 0 \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

Determinar la media, la varianza, la desviación estándar, la mediana, el recorrido intercuartil y el recorrido interdecil de X.

Se deja como ejercicio la gráfica de esta función de densidad de probabilidad y verificar que su integral sea igual a uno. El lector no tendrá ningún problema para notar que esta función de densidad exhibe un rápido decaimiento hacia el eje horizontal; teniendo en cuenta esta propiedad, la distribución anterior puede ser apropiada para representar la edad a la que fallece una persona como resultado de las enfermedades padecidas en su niñez, como la escarlatina y la difteria (hace una generación) y, en mayor frecuencia, la leucemia (en la actualidad).

El valor esperado de X es:

$$E(X) = \frac{1}{8}\int_0^\infty x^{1/2}\exp(-x^{1/2}/4)dx = \frac{1}{8}\int_0^\infty 4u\exp(-u)32udu = 16\Gamma(3) = 32,$$

en donde $u = x^{1/2}/4$, $x = 16u^2$, y $dx = 32udu$. De manera similar:

$$E(X^2) = \frac{1}{8}\int_0^\infty x^{3/2}\exp(-x^{1/2}/4)dx = 256\int_0^\infty u^4\exp(-u)du = 256\Gamma(5) = 6\ 144.$$

de manera tal que $Var(X) = 5\ 120$ y $d.e.(X) = 71.55$.

Para determinar los valores cuantiles, primero se obtendrá la función de distribución acumulativa:

$$F(x) = \frac{1}{8}\int_0^x t^{-1/2}\exp(-t^{1/4}/4)dt = \int_0^{x^{1/2}/4}\exp(-u)du = 1 - \exp(-x^{1/2}/4),$$

en donde $u = t^{1/2}/4$ y $dt = 32udu$. Por definición, la mediana es el valor $x_{0.5}$ tal que $F(x_{0.5}) = 0.5$. Por lo tanto:

$$1 - \exp(-x_{0.5}^{1/2}/4) = 0.5$$

$$\exp(-x_{0.5}^{1/2}/4) = 0.5$$

$$(-x_{0.5}^{1/2}/4) = \ln(0.5)$$

y

$$x_{0.5} = [-4\ln(0.5)]^2 = 7.6872.$$

En otras palabras, el 50% de los valores de X serán menores de 7.6872, a pesar de que la media tiene un valor de 32, lo que constituye una diferencia muy grande entre los valores de la media y la mediana. Para demostrar cuán inapropiada es la media de X como única medida de tendencia central, considérese la probabilidad de que X sea menor que su valor medio:

$$P(X < 32) = F(32) = 1 - \exp(-32^{1/2}/4) = 0.7569.$$

De acuerdo con lo anterior, el valor de 32 para la media difícilmente puede interpretarse como una medida representativa de tendencia central si la probabilidad de que la variable aleatoria exceda el valor de su media es menor de 0.25.

Los percentiles décimo, 25avo, 75avo y 90avo se determinan encontrando el valor de x_q de las ecuaciones $F(x_q) = 0.1,\ 0.25,\ 0.75,$ y 0.90, respectivamente. Por lo tanto:

$$1 - \exp(-x_{0.1}^{1/2}/4) = 0.1$$

$$\exp(-x_{0.1}^{1/2}/4) = 0.9$$

$$x_{0.1} = [-4\ln(0.9)]^2,$$

y $x_{0.1} = 0.1776$. De manera similar, $x_{0.25} = [-4\ln(0.75)]^2 = 1.3242$, $x_{0.75} = [-4\ln(0.25)]^2 = 30.7490$, y $x_{0.9} = [-4\ln(0.1)]^2 = 84.8304$. El recorrido intercuartil de X es $x_{0.75} - x_{0.25} = 30.7490 - 1.3242 = 29.4248$, el recorrido interdecil es $x_{0.9} - x_{0.1} = 84.8304 - 0.1776 = 84.6528$. Nótese que la desviación estándar de X es, aproximadamente 2.5 veces el recorrido intercuartil y casi tan grande como el recorrido interdecil. Este resultado, junto con los hechos de que el 25% de los valores son menores de 1.3242, el 50% es menor de 7.6872 y el 75% menores de 30.49, demuestran que la varianza, y por lo tanto la desviación estándar, son inadecuadas como únicas medidas de variabilidad.

3.7 Funciones generadoras de momentos

Hasta este momento se han presentado distintas formas para determinar los momentos de una variable aleatoria dada su distribución de probabilidad. Como método alternativo se presenta la esperanza de cierta función conocida como *función generadora de momentos.*

Definición 3.14 Sea X una variable aleatoria. El valor esperado de *exp(tX)* recibe el nombre de función generadora de momentos, y se denota por $m_X(t)$, si el valor es-

perado existe para cualquier valor de t en algún intervalo $-c < t < c$ en donde c es un número positivo. En otras palabras:

$$m_X(t) = E[\exp(tX)] = \sum_x \exp(tx)p(x) \qquad \text{si } X \text{ es discreta, o}$$

$$m_X(t) = E[\exp(tX)] = \int_{-\infty}^{\infty} \exp(tx)f(x)dx \qquad \text{si } X \text{ es continua.}$$

Nótese que $m_X(t)$ nada más es función del argumento t. Si $t = 0$, entonces $m_X(0) = E(e^0) = 1$. Si la función generadora de momentos existe, puede demostrarse que es única y que determina por completo la distribución de probabilidad de X. En otras palabras, si dos variables aleatorias tienen la misma función generadora de momentos, entonces tienen la misma distribución de probabilidad. Este resultado se utilizará, de manera extensa, en el capítulo siete.

Si la función generadora de momentos existe para $-c < t < c$, entonces existen las derivadas de ésta de todas las órdenes para $t = 0$. Lo anterior asegura que $m_X(t)$ generará todos los momentos de X alrededor del origen. Para demostrar lo anterior, se diferencia $m_X(t)$ con respecto a t, y se evalúa la derivada en $t = 0$. Suponiendo que pueden intercambiarse los símbolos de diferenciación y esperanza, se tiene:

$$\begin{aligned} \left.\frac{dm_X(t)}{dt}\right|_{t=0} &= \left.\frac{d}{dt} E[\exp(tX)]\right|_{t=0} \\ &= E\left\{\frac{d}{dt}[\exp(tX)]\right\} \\ &= E[X\exp(tX)]|_{t=0} \\ &= E(X) = \mu. \end{aligned}$$

Al tomar la segunda derivada y evaluar en $t = 0$.

$$\begin{aligned} \left.\frac{d^2m_X(t)}{dt^2}\right|_{t=0} &= \left.\frac{d^2}{dt^2} E[\exp(tX)]\right|_{t=0} \\ &= E\left\{\frac{d^2}{dt^2}[\exp(tX)]\right\} \\ &= E\left\{\frac{d}{dt}[X\exp(tX)]\right\} \\ &= E[X^2\exp(tX)]|_{t=0} \\ &= E(X^2) = \mu_2'. \end{aligned}$$

Al continuar este proceso de diferenciación se puede deducir que se obtiene el

$$\left.\frac{d^r m_X(t)}{dt^r}\right|_{t=0} = \left.\frac{d^r}{dt^r} E[\exp(tX)]\right|_{t=0}$$

$$= E\left\{\frac{d^r}{dt^r}[\exp(tX)]\right\}$$

$$= E[X^r \exp(tX)]\,|_{t=0}$$

$$= E(X^r) = \mu'_r.$$

mismo resultado si se reemplaza la función exponencial por su expansión en serie de potencias

$$E[\exp(tX)] = E\left(1 + tX + \frac{t^2X^2}{2!} + \cdots + \frac{t^rX^r}{r!} + \cdots\right)$$

y se toman las derivadas con respecto a t, evaluando cada una de éstas en $t = 0$.

La noción de una función generadora de momentos puede extenderse a otros puntos de referencia, además del origen. En particular, se define una función central generadora de momentos la que, si existe, generará todos los momentos centrales de una distribución de probabilidad.

Definición 3.15 Sea X una variable aleatoria. El valor esperado de $\exp[t(X - \mu)]$ recibe el nombre de función generadora de momentos central y denota por $m_{X-\mu}(t)$, si el valor esperado existe para cualquier t en algún intervalo $-c < t < c$ en donde c es un número positivo.

$$m_{X-\mu}(t) = E\{\exp[t(X - \mu)]\} = \sum_x \exp[t(x - \mu)]p(x) \qquad \text{si } X \text{ es discreta, o}$$

$$m_{X-\mu}(t) = E\{\exp[t(X - \mu)]\} = \int_{-\infty}^{\infty} \exp[t(x - \mu)]f(x)dx \qquad \text{si } X \text{ es continua.}$$

La comprobación de que $m_{X-\mu}(t)$ genera todos los momentos centrales se deja como ejercicio al lector.

Ejemplo 3.13 Sea X una variable aleatoria con función de densidad de probabilidad

$$f(x) = \begin{cases} \dfrac{1}{\theta}\exp(-x/\theta) & x > 0, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

en donde θ es un número mayor que cero. Determinar la función generadora de momentos de X.

Por definición

$$m_X(t) = \frac{1}{\theta}\int_0^{\infty} \exp(tx)\exp(-x/\theta)dx$$

$$= \frac{1}{\theta}\int_0^{\infty} \exp\left[-x\left(\frac{1}{\theta} - t\right)\right]dx$$

$$= -\frac{\theta}{\theta(1-\theta t)}\exp\left[-x\left(\frac{1}{\theta} - t\right)\right]\Bigg|_0^{\infty}$$

$$= (1-\theta t)^{-1}.$$

Por lo tanto:

$$\left.\frac{dm_X(t)}{dt}\right|_{t=0} = \theta(1-\theta t)^{-2}\Big|_{t=0}$$

$$= \theta = E(X),$$

y

$$\left.\frac{d^2m_X(t)}{dt^2}\right|_{t=0} = 2\theta^2(1-\theta t)^{-3}\Big|_{t=0}$$

$$= 2\theta^2 = E(X^2).$$

dando como resultado, $Var(X) = 2\theta^2 - \theta^2 = \theta^2$, y así sucesivamente.

Ejemplo 3.14 Sea X una variable aleatoria discreta con función de probabilidad:

$$p(x) = \frac{\exp(-\lambda)\lambda^x}{x!} \qquad x = 0, 1, 2, \ldots,$$

en donde λ es un número mayor que cero. Determinar la función generadora de momentos de X.

De acuerdo con la definición se tiene:

$$m_X(t) = \sum_{x=0}^{\infty} \frac{\exp(tx)\exp(-\lambda)\lambda^x}{x!}$$

$$= \exp(-\lambda)\sum_{x=0}^{\infty} \frac{[\lambda\exp(t)]^x}{x!}.$$

Dado que:

$$\sum_{x=0}^{\infty} \frac{[\lambda\exp(t)]^x}{x!} = 1 + \lambda e^t + \frac{\lambda^2 e^{2t}}{2!} + \cdots + \frac{\lambda^r e^{rt}}{r!} + \cdots$$

$$= \exp[\lambda\exp(t)],$$

entonces:

$$m_X(t) = \exp(-\lambda)\exp[\lambda \exp(t)].$$

Por lo tanto:

$$\left.\frac{dm_X(t)}{dt}\right|_{t=0} = \lambda \exp(-\lambda)\exp(t)\exp\left[\lambda \exp(t)\right]\Big|_{t=0}$$

$$= \lambda = E(X).$$

Referencias

1. J. G. Freund, *Mathematical statistics*, 2nd ed., Prentice-Hall, Englewood Cliffs, N.J., 1971.
2. P. G. Hoel, *Introduction to mathematical statistics*, 4th ed., Wiley, New York, 1971.
3. W. Mendenhall and R. L. Schaeffer, *Mathematical statistics with applications*, Duxbury, North Scituate, Mass., 1973.

Ejercicios

3.1. Sea X una variable aleatoria que representa el número de llamadas telefónicas que recibe un conmutador en un intervalo de cinco minutos y cuya función de probabilidad está dada por $p(x) = e^{-3}\,(3)^x/x!$, $x = 0, 1, 2, \ldots$.

a) Determinar las probabilidades de que X sea igual a 0, 1, 2, 3, 4, 5, 6 y 7.
b) Graficar la función de probabilidad para estos valores de X.
c) Determinar la función de distribución acumulativa para estos valores de X.
d) Graficar la función de distribución acumulativa.

3.2. Sea X una variable aleatoria discreta. Determinar el valor de k para que la función $p(x) = k/x$, $x = 1, 2, 3, 4$, sea la función de probabilidad de X. Determinar $P(1 \leq X \leq 3)$.

3.3. Sea X una variable aleatoria continua.

a) Determinar el valor de k, de manera tal que la función

$$f(x) = \begin{cases} kx^2 & -1 \leq x \leq 1, \\ 0 & \text{para cualquier otro valor} \end{cases}$$

sea la función de densidad de probabilidad de X.
b) Determinar la función de distribución acumulativa de X y graficar $F(x)$.
c) Calcular $P(X \geq 1/2)$ y $P(-1/2 \leq X \leq 1/2)$.

3.4. Sea X una variable aleatoria continua.

a) Determinar el valor de k para que la función

$$f(x) = \begin{cases} k\exp(-x/5) & x > 0, \\ 0 & \text{para cualquier otro valor} \end{cases}$$

sea la función de densidad de probabilidad de X

b) Graficar $f(x)$.
c) Calcular $P(X \leq 5)$ y $P(0 \leq X \leq 8)$.
d) Determinar $F(x)$ y graficarla.

3.5. La duración en horas de un componente electrónico, es una variable aleatoria cuya función de distribución acumulativa es $F(x) = 1 - \exp(-x/100)$, $x > 0$.

a) Determinar la función de probabilidad de X.
b) Determinar la probabilidad de que el componente trabaje más de 200 horas.

3.6. La función de distribución acumulativa de una variable aleatoria está dada por

$$F(x) = \begin{cases} 0 & x < 0, \\ 2x - x^2 & 0 < x < 1, \\ 1 & x > 1. \end{cases}$$

a) Graficar $F(x)$.
b) Obtener $P(X < 1/2)$ y $P(X > 3/4)$.
c) Determinar $f(x)$.

3.7. Sea X una variable aleatoria que representa el número de clientes que llega a una tienda en un periodo de una hora. Dada la siguiente información

x	0	1	2	3	4	5	6	7	8
$p(x)$	0.05	0.10	0.10	0.10	0.20	0.25	0.10	0.05	0.05

encontrar $E(X)$ y $Var(X)$.

3.8. Una compañía de seguros debe determinar la cuota anual a cobrarse por un seguro de $50 mil para hombres cuya edad se encuentra entre los 30 y 35 años. Con base en las tablas actuariales el número de fallecimientos al año, para este grupo, es de 5 por cada mil. Si X es la variable aleatoria que representa la ganancia de la compañía de seguros, determinar el monto de la cuota anual para que la compañía no pierda, a pesar de tener un número grande de tales seguros.

3.9. La función de densidad de probabilidad de una variable aleatoria X está dada por:

$$f(x) = \begin{cases} 2(1 - x) & 0 < x < 1, \\ 0 & \text{para cualquier otro valor} \end{cases}$$

Determinar:

a) $E(X)$ *b)* $Var(X)$

3.10. Sea X una variable aleatoria que representa la magnitud de la desviación, a partir de un valor prescrito, del peso neto de ciertos recipientes, los que se llenan mediante una máquina. La función de densidad de probabilidad de X está dada por:

$$f(x) = \begin{cases} 1/10 & 0 < x < 10 \\ 0 & \text{para cualquier otro valor} \end{cases}$$

Determinar:

a) $E(X)$ *c)* $\alpha_3(X)$
b) $Var(X)$ *d)* $\alpha_4(X)$

3.11. Supóngase que la duración en minutos de una llamada de negocios, es una variable aleatoria cuya función de densidad de probabilidad está determinada

$$f(x) = \begin{cases} \frac{1}{4}\exp(-x/4) & x > 0, \\ 0 & \text{para cualquier otro valor} \end{cases}$$

Determinar:

a) $E(X)$
b) $Var(X)$
c) $\alpha_3(X)$
d) $\alpha_4(X)$
e) Refiérase al ejercicio 3.10. Basándose en sus respuestas a las preguntas *a*, a *d* del problema 3.11, compare las dos distribuciones de probabilidades. ¿Cuál muestra la mayor dispersión relativa?

3.12. La calificación promedio en una prueba de estadística fue de 62.5 con una desviación estándar de 10. El profesor sospecha que el examen fue difícil. De acuerdo con lo anterior, desea ajustar las calificaciones de manera que el promedio sea de 70 y la desviación estándar de 8. ¿Qué ajuste del tipo $aX + b$, debe utilizar?

3.13. Sea X una variable aleatoria con media μ y varianza σ^2.

a) Evaluar $E(X - c)^2$ en términos de μ y σ^2 en donde c es una constante.
b) ¿Para qué valor de c es $E(X - c)^2$ mínimo?

3.14. Con respecto al ejercicio 3.11, demostrar que la variable aleatoria $Y = (X - 4)/4$ tiene media cero y desviación estándar uno. Demostrar que los factores de forma, primero y segundo, de la distribución de Y son los mismos de la distribución de X.

3.15. Considérese la función de densidad de probabilidad de X dada en el ejercicio 3.9. Determinar la desviación media de X y compararla con su desviación estándar.

3.16. Considérese la función de densidad de probabilidad de X dada en el ejercicio 3.10. Determinar la desviación media de X y compararla con su desviación estándar.

3.17. Supóngase que el ingreso semanal de un asesor profesional es una variable aleatoria cuya función de densidad de probabilidad está determinada por:

$$f(x) = \begin{cases} \frac{1}{800}\exp(-x/800) & x > 0, \\ 0 & \text{para cualquier otro valor} \end{cases}$$

a) Determinar los ingresos medios y medianos.
b) Determinar el recorrido intercuartil.
c) Determinar el recorrido interdecil.
d) Determinar la probabilidad de que el ingreso semanal exceda al ingreso promedio.

3.18. Comprobar que la función generadora de momentos central de una variable aleatoria X, genera todos los momentos centrales de X.

3.19. La función de densidad de probabilidad de una variable aleatoria X está determinada:

$$f(x) = \begin{cases} \frac{1}{16}x\exp(-x/4) & x > 0, \\ 0 & \text{para cualquier otro valor} \end{cases}$$

a) Determinar la función generadora de momentos de X.
b) Utilizar la función generadora de momentos para encontrar la media y la varianza de X.

3.20. Considérese la función de densidad de probabilidad dada en el ejercicio 3.11. Encontrar la función generadora de momentos y utilizarla para comprobar los valores de la media y la varianza, determinados en el ejercicio 3.11.

3.21. Sea X una variable aleatoria discreta con función de probabilidad $p(x)$ $x = 0,1,2,\ldots,n$, y sean a, b, y c constantes. Demostrar que $E(c) = c$, $E(aX + b) = aE(X) + b$, y $E[g(X) + h(X)] = E[g(X)] + E[h(X)]$, en donde $g(x)$ y $h(x)$ son funciones de X.

3.22. Para la variable aleatoria discreta del ejercicio anterior, utilizar las definiciones 3.8 y 3.9 para demostrar que $Var(X) = E(X^2) - E^2(X)$.

CAPÍTULO CUATRO

Algunas distribuciones discretas de probabilidad

4.1 Introducción

En el capítulo dos se establecieron algunos principios básicos de probabilidad. En el capítulo tres estos principios se aplicaron para definir variables aleatorias y distribuciones de probabilidad así como para desarrollar sus propiedades generales. En los capítulos cuatro y cinco se examinarán con detalle algunas distribuciones específicas de probabilidad que han demostrado, empíricamente, ser modelos útiles para diversos problemas prácticos. A pesar de ello tales distribuciones presentan un carácter teórico en el sentido en que sus funciones de probabilidad o de densidad de probabilidad se deducen matemáticamente con base en ciertas hipótesis que se suponen válidas para los fenómenos aleatorios. La elección de una distribución de probabilidad para representar un fenómeno de interés práctico debe estar motivada tanto por la comprensión de la naturaleza del fenómeno en sí, como por la posible verificación de la distribución seleccionada a través de la evidencia empírica. En todo momento debe evitarse aceptar de manera tácita una determinada distribución de probabilidad como modelo de un problema práctico.

Se examinarán varias distribuciones tanto discretas como continuas. En cada caso se expondrán detalladamente las características distintivas de las distribuciones particulares de probabilidad y se deducirán o se establecerán sus medias, varianzas, factores de forma, y otras medidas descriptivas numéricas. Como se sugirió en el capítulo uno, una distribución de probabilidad está caracterizada, de manera general, por una o más cantidades que reciben el nombre de parámetros de la distribución. Un parámetro puede tomar cualquier valor de un conjunto dado y, en ese sentido, define una familia de distribuciones de probabilidad, que tendrán la misma función genérica de probabilidad o función de densidad de probabilidad. Se tratarán varios tipos de parámetros tales como el conteo, la proporción, la rapidez, la localización y la forma. Se adoptarán las letras n y k para referirse a los parámetros de conteo, p para la proporción λ para la rapidez, μ para la localización, σ y θ para la escala, y α y β para la forma. Cuando la presentación sea de una naturaleza muy

general y no se esté tratandoningún tipo de parámetro en particular, se empleará θ para designar ese parámetro.

Los parámetros de conteo y de proporción son autoexplicatorios. Un parámetro de rapidez representa la rapidez en que ocurre un evento aleatorio en el tiempo o en el espacio. Un parámetro de localización relaciona la función (densidad) de probabilidad con el origen de la escala de medición, localizándola sobre el eje de las x sin tener algún efecto sobre su apariencia. La presencia de un parámetro de localización μ en la función de probabilidad es siempre de la forma $(x - \mu)$. Un parámetro de escala es una cantidad que relaciona las unidades físicas de la variable aleatoria y de esta forma la escala. Un parámetro de escala influye sobre la dispersión de una variable aleatoria, y de esta forma afecta la apariencia de la función de probabilidad. La aparición de un parámetro de escala en la función de probabilidad es de la forma x/θ. Un parámetro de forma afecta la forma de la función de probabilidad en diverso grado, dependiendo del modelo en particular. A pesar de que en muchas ocasiones el parámetro de forma se encuentra en un exponente en la función de probabilidad, no existe ninguna forma estándar en la que pueda asociarse a x sin importar su aparición en la función de probabilidad.

Se examinarán con detalle cuatro familias de distribuciones de probabilidad discreta y se harán comentarios sobre su aplicación. Estas son las distribuciones binomial, Poisson, hipergeométrica y la binomial negativa.

4.2 La distribución binomial

Es una de las distribuciones discretas de probabilidad más útiles. Sus áreas de aplicación incluyen inspección de calidad, ventas, mercadotecnia, medicina, investigación de opiniones y otras. Se puede imaginar un experimento en el que el resultado es la ocurrencia o la no ocurrencia de un evento. Sin pérdida de generalidad, llámese "éxito" a la ocurrencia del evento y "fracaso" a su no ocurrencia. Además, sea p la probabilidad de éxito cada vez que el experimento se lleva a cabo y $1 - p$ la probabilidad de fracaso. Supóngase que el experimento se realiza n veces, y cada uno de éstos es independiente de todos los demás, y sea X la variable aleatoria que representa el número de éxitos en los n ensayos. El interés está en determinar la probabilidad de obtener exactamente $X = x$ éxitos durante los n ensayos. Las dos suposiciones claves para la distribución binomial son:

1. La probabilidad de éxito p permanece constante para cada ensayo.
2. Los n ensayos son independientes entre sí.

Varios problemas prácticos parecen adherirse razonablemente a las suposiciones anteriores. Por ejemplo, un proceso de manufactura produce un determinado producto en el que algunas unidades se encuentran defectuosas. Si la proporción de unidades defectuosas producidas por este proceso es constante durante un periodo razonable y, si como procedimiento de rutina, se seleccionan aleatoriamente un determinado número de unidades, entonces las proposiciones de probabilidad con respecto al número de artículos defectuosos puede hacerse mediante el empleo de la distribución binomial. La publicidad para la venta de un producto también puede considerarse otro ejemplo.

Si se supone que la probabilidad de venta es constante para todas las personas, la distribución binomial será el modelo de probabilidad adecuado puesto que las personas tienen un criterio independiente para comprar. Como ejemplo final, el Centro para el Control de Enfermedades tiene, entre sus distintas funciones, la responsabilidad de vigilar las enfermedades transmisibles. Para cumplir con ella, debe examinar la propagación de una enfermedad determinada con base en la probabilidad. Es dudoso que la probabilidad de contraer una enfermedad transmisible, sea constante para toda la población. Sin embargo, para una parte de ésta, por ejemplo las personas que tienen una edad determinada, sí puede ser constante, de manera tal que la distribución binomial puede ser un modelo de probabilidad adecuado.

Para obtener la función de probabilidad de la distribución binomial, primero se determina la probabilidad de tener, en n ensayos, x éxitos consecutivos seguidos de $n-x$ fracasos consecutivos. Dado que, por hipótesis, los n ensayos son independientes de la definición 2.15, se tiene:

$$\underbrace{p \cdot p \cdots p}_{x \text{ términos}} \cdot \underbrace{(1-p)(1-p)\cdots(1-p)}_{(n-x) \text{ términos}} = p^x(1-p)^{n-x}.$$

La probabilidad de obtener exactamente x éxitos y $n-x$ fracasos en cualquier otro orden es la misma puesto que los factores p y $(1-p)$ se reordenan de acuerdo con el orden particular. Por lo tanto, la probabilidad de tener x éxitos y $n-x$ fracasos en cualquier orden, es el producto de $p^x(1-p)^{n-x}$ por el número de órdenes distintos. Este último es el número de combinaciones de n objetos tomando x a la vez. De acuerdo con lo anterior se tiene la siguiente definición:

Definición 4.1 Sea X una variable aleatoria que representa el número de éxitos en n ensayos y p la probabilidad de éxito con cualquiera de éstos. Se dice entonces que X tiene una distribución binomial con función de probabilidad.*

$$p(x; n, p) = \begin{cases} \dfrac{n!}{(n-x)!\,x!}\, p^x(1-p)^{n-x} & x = 0, 1, 2, \ldots, n, \\ 0 \text{ para cualquier otro valor.} & 0 \leq p \leq 1, \quad \text{para } n \text{ entero.} \end{cases} \tag{4.1}$$

Los parámetros de la distribución binomial son n y p. Éstos definen una familia de distribuciones binomiales, en donde cada miembro tiene la función de probabilidad determinada por (4.1). Para ilustrar el efecto de estos parámetros, la figura 4.1 proporciona algunas gráficas de la distribución binomial. Se dará más información sobre éstas cuando se discutan los momentos y otras medidas descriptivas.

El nombre "distribución binomial" proviene del hecho de que los valores de $p(x; n, p)$ para $x = 0, 1, 2 \ldots n$ son los términos sucesivos de la expansión binomial de $[(1-p) + p]^n$; esto es,

* Para mantener la consistencia, se empleará la notación $p(\;)$ para indicar la función básica de probabilidad. El autor no piensa que el lector se confundirá por el empleo de $p(x; n, p)$ para la función de probabilidad binomial y el uso de la letra p para el parámetro de proporción.

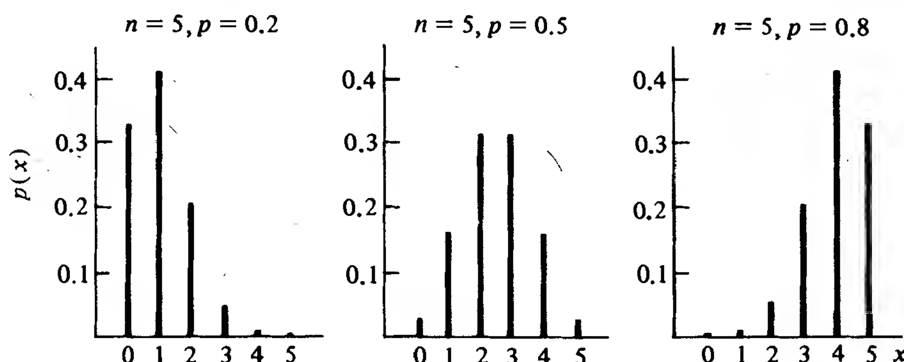


FIGURA 4.1 Gráficas de la función binomial de probabilidad

$$[(1 - p) + p]^n = (1 - p)^n + n(1 - p)^{n-1}p + \frac{n(n-1)}{2!}(1 - p)^{n-2}p^2 + \cdots + p^n$$

$$= \sum_{x=0}^{n} \frac{n!}{(n-x)!x!} p^x(1 - p)^{n-x}$$

$$= \sum_{x=0}^{n} p(x; n, p).$$

Pero dado que $[(1 - p) + p]^n = 1$ y $p(x; n, p) \geqslant 0$ para $x = 0, 1, 2 \ldots n$, este hecho también verifica que $p(x; n, p)$ es una función de probabilidad.

Para ilustrar el cálculo de probabilidad mediante el empleo de (4.1) sea $n = 5$ y $p = 0.4$ entonces:

$$p(x; 5, 0.4) = \frac{5!}{(5-x)!x!}(0.4)^x(0.6)^{5-x}, \quad x = 0, 1, 2, 3, 4, 5;$$

así:

$$p(0; 5, 0.4) = \frac{5!}{(5-0)!0!}(0.4)^0(0.6)^{5-0} = 0.0778,$$

$$p(1; 5, 0.4) = \frac{5!}{(5-1)!1!}(0.4)^1(0.6)^{5-1} = 0.2592,$$

$$p(2; 5, 0.4) = \frac{5!}{(5-2)!2!}(0.4)^2(0.6)^{5-2} = 0.3456,$$

$$p(3; 5, 0.4) = \frac{5!}{(5-3)!3!}(0.4)^3(0.6)^{5-3} = 0.2304,$$

$$p(4; 5, 0.4) = \frac{5!}{(5-4)!4!}(0.4)^4(0.6)^{5-4} = 0.0768,$$

$$p(5; 5, 0.4) = \frac{5!}{(5-5)!5!}(0.4)^5(0.6)^{5-5} = 0.0102.$$

La probabilidad de que una variable aleatoria X sea menor o igual a un valor específico de x, se determina por la función de distribución acumulativa

$$P(X \leqslant x) = F(x; n, p) = \sum_{i=0}^{x} \binom{n}{i} p^i(1-p)^{n-i}. \qquad (4.2)$$

La distribución binomial se ha tabulado de manera extensa para distintos valores de n y p, ya sea mediante el empleo de (4.1) o (4.2) o ambas. En la tabla A del apéndice, se proporcionan las probabilidades acumulativas para distintos valores de x, n, y p. Pueden determinarse las probabilidades individuales mediante el empleo de esta tabla puesto que la variable aleatoria binomial tiene un valor entero, y la propiedad

$$p(x; n, p) = F(x; n, p) - F(x-1; n, p)$$

se verifica. Para ilustrar el uso de la tabla A, sea $n = 10$ y $p = 0.3$. La probabilidad de que X pueda ser cuatro es:

$$P(X \leqslant 4) = F(4; 10, 0.3) = 0.8497;$$

la probabilidad de que X sea mayor de dos es:

$$P(X > 2) = P(X \geqslant 3) = 1 - P(X \leqslant 2) = 1 - F(2; 10, 0.3) = 0.6172;$$

y la probabilidad de que X sea de exactamente cinco es:

$$p(5; 10, 0.3) = F(5; 10, 0.3) - F(4; 10, 0.3) = 0.1030.$$

Debe notarse que si $n = 1$, la función de probabilidad binomial se reduce a:

$$p(x; p) = \begin{cases} p^x(1-p)^{1-x} & x = 0, 1, \\ 0 & \text{para cualquier otro valor,} \end{cases} \qquad (4.3)$$

que es la función de probabilidad de la distribución puntual o de Bernoulli. La distribución de Bernoulli recibe este nombre por el probabilista suizo Jacques Bernoulli (1654-1705) quien desarrolló por primera vez el concepto de ensayos independientes.

Ejemplo 4.1 Todos los días se seleccionan, de manera aleatoria, 15 unidades de un proceso de manufactura con el propósito de verificar el porcentaje de unidades defectuosas en la producción. Con base en información pasada, la probabilidad de tener una unidad defectuosa es de 0.05. La gerencia ha decidido detener la producción

cada vez que una muestra de 15 unidades tenga dos o más defectuosas. ¿Cuál es la probabilidad de que, en cualquier día, la producción se detenga?

Si el modelo apropiado para esta situación es la distribución binomial, se puede suponer que las 15 unidades que se seleccionan al día, constituyen un conjunto de ensayos independientes de manera tal que la probabilidad de tener una unidad defectuosa es 0.05 entre ensayos. Sea X el número de unidades defectuosas que se encuentran entre las 15. Para $n = 15$ y $p = 0.05$, la probabilidad de que la producción se detenga es igual a la probabilidad de que X sea igual o mayor que dos. De esta manera:

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - F(1; 15, 0.05) = 0.1709.$$

Ejemplo 4.2 Supóngase que para personas de determinada edad, la probabilidad de que mueran por una enfermedad transmisible es 0.001. ¿Cuántas personas de este grupo pueden exponerse a la enfermedad de manera que la probabilidad de que no más de una persona muera sea por lo menos 0.95?

Para aplicar la distribución binomial a esta situación, la suposición crucial es que la probabilidad de muerte es constante para todos los individuos que forman parte del grupo y que contraen la enfermedad. Sea X el número de muertes que ocurren en n individuos por haber contraído el padecimiento. El valor de n para que la probabilidad de que X sea menor o igual a uno tenga un valor mayor o igual a 0.95:

$$P(X \leq 1) = F(1; n, 0.001) \geq 0.95,$$

y para la igualdad:

$$\sum_{x=0}^{1} \binom{n}{x} (0.001)^x (0.999)^{n-x} = 0.95$$

$$\binom{n}{0} (0.001)^0 (0.999)^n + \binom{n}{1} (0.001)^1 (0.999)^{n-1} = 0.95$$

$$(0.999)^{n-1}(0.999 + 0.001n) = 0.95.$$

Esta ecuación no se resuelve de manera explícita para n; sin embargo, mediante el empleo de técnicas iterativas* puede determinarse que el valor entero de n que satisface la ecuación es $n = 356$.

En este punto se determinarán los momentos para la distribución binomial. Se ilustrarán tanto el método directo, con base en la definición 3.8, como el método indirecto, con base en la función generadora de momentos.

* Una técnica iterativa es un método numérico para resolver una ecuación mediante una sucesión de valores hasta que el último valor se encuentra muy cercano al que satisface la ecuación.

Por la definición 3.8, el primer momento alrededor del cero de la variable aleatoria binomial X es el valor esperado de X,

$$E(X) = \sum_{x=0}^{n} x \frac{n!}{(n-x)!x!} p^x(1-p)^{n-x}$$

$$= \sum_{x=1}^{n} x \frac{n!}{(n-x)!x!} p^x(1-p)^{n-x}$$

$$= \sum_{x=1}^{n} \frac{n!}{(n-x)!(x-1)!} p^x(1-p)^{n-x},$$

en donde se ha escrito la suma desde uno hasta n, dado que cuando $x = 0$ el primer término es cero y se cancela la x del numerador con la x en $x!$. Factorizando n y p, se tiene:

$$E(X) = np \sum_{x=1}^{n} \frac{(n-1)!}{(n-x)!(x-1)!} p^{x-1}(1-p)^{n-x}$$

Si $y = x - 1$ y $m = n - 1$, entonces:

$$E(X) = np \sum_{y=0}^{m} \frac{m!}{(m-y)!y!} p^y(1-p)^{n-y}.$$

Pero $p(y;\ m,\ p) = [m!/(m-y)!y!]p^y(1-p)^{m-y}$ es la función de probabilidad de una variable aleatoria binomial Y con parámetros $m = n - 1$ y p; de esta manera $\Sigma_{y=0}^{m}\ p(y;\ m,\ p) = 1$, y la media de una variable aleatoria binomial es:

$$E(X) = \mu = np. \tag{4.4}$$

Para obtener la varianza, se necesita el segundo momento alrededor del cero, μ_2', o:

$$E(X^2) = \sum_{x=0}^{n} x^2 p(x; n, p);$$

pero, en el término $x^2/x!$ se cancelará una sola x en el numerador, y la que resta evitará que la suma se manipule de la misma forma en que se determinó la media. La alternativa es escribir x^2 como:

$$x^2 = x(x-1) + x;$$

de esta manera se tiene:

$$E(X^2) = E[X(X-1)] + E(X). \tag{4.5}$$

Dado que $E(X)$ ya se ha determinado, puede usarse el mismo procedimiento para

evaluar $E[X(X-1)]$:

$$E[X(X-1)] = \sum_{x=0}^{n} x(x-1)\frac{n!}{(n-x)!x!}p^x(1-p)^{n-x}$$

$$= \sum_{x=2}^{n} x(x-1)\frac{n!}{(n-x)!x!}p^x(1-p)^{n-x}$$

$$= \sum_{x=2}^{n} \frac{n!}{(n-x)!(x-2)!}p^x(1-p)^{n-x}$$

$$= n(n-1)p^2\sum_{x=2}^{n} \frac{(n-2)!}{(n-x)!(x-2)!}p^{x-2}(1-p)^{n-x}$$

Nótese que en los pasos anteriores se escribió la suma a partir de dos porque los dos primeros términos son cero, se canceló $x(x-1)$, y se factorizó $n(n-1)p^2$. Sea $y = x - 2$ y $m = n - 2$; entonces:

$$E[X(X-1)] = n(n-1)p^2\sum_{y=0}^{m} \frac{m!}{(m-y)!y!}p^y(1-p)^{m-y}$$

$$= n(n-1)p^2\sum_{y=0}^{m} p(y; m, p)$$

$$= n(n-1)p^2.$$

De (4.5)

$$E(X^2) = \mu_2' = n(n-1)p^2 + np.$$

De esta manera, la varianza de una variable aleatoria binomial es:

$$Var(X) = \mu_2' - \mu^2$$

$$= n(n-1)p^2 + np - n^2p^2$$

$$= np\,[(n-1)p + 1 - np]$$

$$= np(1-p). \tag{4.6}$$

Este método general puede extenderse para determinar los momentos de orden superior. Por ejemplo, para obtener el tercer momento alrededor del cero, se determina $E[X(X-1)(X-2)]$ dado que:

$$E[X(X-1)(X-2)]^* = \mu_3' - 3\mu_2' + 2\mu. \tag{4.7}$$

De manera similar, para el cuarto momento alrededor del cero se evalúa $E[X(X-1)$

$(X - 2)(X - 3)]$ dado que:

$$E[X(X - 1)(X - 2)(X - 3)]^* = \mu_4' - 6\mu_3' + 11\mu_2' - 6\mu. \tag{4.8}$$

Para una variable aleatoria binomial:

$$\begin{aligned} E[X(X - 1)(X - 2)] &= \sum_{x=0}^{n} x(x - 1)(x - 2) \frac{n!}{(n - x)!x!} p^x(1 - p)^{n-x} \\ &= \sum_{x=3}^{n} \frac{n!}{(n - x)!(x - 3)!} p^x(1 - p)^{n-x} \\ &= n(n - 1)(n - 2)p^3 \sum_{x=3}^{n} \frac{(n - 3)!}{(n - x)!(x - 3)!} p^{x-3}(1 - p)^{n-x} \\ &= n(n - 1)(n - 2)p^3 \sum_{y=0}^{m} \frac{m!}{(m - y)!y!} p^y(1 - p)^{m-y} \\ &= n(n - 1)(n - 2)p^3. \end{aligned}$$

Mediante el empleo de (4.7),

$$\begin{aligned} \mu_3' - 3\mu_2' + 2\mu &= n(n - 1)(n - 2)p^3 \\ \mu_3' &= n(n - 1)(n - 2)p^3 + 3[n(n - 1)p^2 + np] - 2np \\ &= n(n - 1)(n - 2)p^3 + 3n(n - 1)p^2 + np. \end{aligned} \tag{4.9}$$

El tercer momento central μ_3 puede determinarse por (3.8),

$$\mu_3 = n(n - 1)(n - 2)p^3 + 3n(n - 1)p^2 + np - 3np[n(n - 1)p^2 + np] + 2n^3p^3,$$

la que, después de un poco de álgebra, se reduce a:

$$\mu_3 = np(1 - p)(1 - 2p). \tag{4.10}$$

Por lo tanto, de (3.9) el tercer momento estandarizado de la distribución binomial es:

$$\begin{aligned} \alpha_3 &= \frac{np(1 - p)(1 - 2p)}{[np(1 - p)]^{3/2}} \\ &= \frac{np(1 - p)(1 - 2p)}{np(1 - p)[np(1 - p)]^{1/2}} \\ &= \frac{1 - 2p}{[np(1 - p)]^{1/2}}. \end{aligned} \tag{4.11}$$

* Expresiones como éstas dan lo que se conoce como momentos factoriales. De hecho, el r-ésimo momento factorial de una variable aleatoria X es $E[X(X - 1)(X - 2) \cdots (X - r + 1)]$.

Para el cuarto momento alrededor del cero, se tiene:

$$E[X(X-1)(X-2)(X-3)] = \sum_{x=0}^{n} x(x-1)(x-2)(x-3) \cdot \frac{n!}{(n-x)!x!} p^x(1-p)^{n-x}$$

$$= \sum_{x=4}^{n} \frac{n!}{(n-x)!(x-4)!} p^x(1-p)^{n-x}$$

$$= n(n-1)(n-2)(n-3)p^4 \cdot \sum_{x=4}^{n} \frac{(n-4)!}{(n-x)!(x-4)!} p^{x-4}(1-p)^{n-x}$$

$$= n(n-1)(n-2)(n-3)p^4 \cdot \sum_{y=0}^{m} \frac{m!}{(m-y)!y!} p^y(1-p)^{m-y}$$

$$= n(n-1)(n-2)(n-3)p^4.$$

Sustituir en (4.8) y para resolver μ_4', se tiene:

$$\mu_4' = n(n-1)(n-2)(n-3)p^4 + 6[n(n-1)(n-2)p^2 + 3n(n-1)p^2 + np] - 11[n(n-1)p^2 + np] + 6np. \quad (4.12)$$

De acuerdo con (3.10), el cuarto momento central es:

$$\mu_4 = \mu_4' - 4\mu\mu_3' + 6\mu^2\mu_2' - 3\mu^4,$$

el que, después de una sustitución adecuada y un poco de manipulación algebraica, es

$$\mu_4 = np(1-p)\{3np(1-p) + [1 - 6p(1-p)]\}. \quad (4.13)$$

De acuerdo con (3.11), el cuarto momento estandarizado de la distribución binomial es:

$$\alpha_4 = \frac{np(1-p)\{3np(1-p) + [1 - 6p(1-p)]\}}{n^2p^2(1-p)^2} = 3 + \frac{[1 - 6p(1-p)]}{np(1-p)}. \quad (4.14)$$

Las propiedades básicas de la distribución binomial se encuentran resumidas en la tabla 4.1. Nótese que la media de una variable aleatoria binomial es el producto del número de ensayos y la probabilidad de éxito en cada uno de éstos y la varianza es el producto de la media por la probabilidad de tener un fracaso. La varianza de una variable aleatoria binomial siempre es menor que el valor de su media.

TABLA 4.1 Propiedades básicas de la distribución binomial.

Función de probabilidad			*Parámetros*
$p(x; n, p) = \dfrac{n!}{(n-x)!x!} p^x(1-p)^{n-x}$ $x = 0, 1, 2, \ldots, n$			n, entero positivo p, $0 \leq p \leq 1$
Media	*Varianza*	*Coeficiente de sesgo*	*Curtosis relativa*
np	$np(1-p)$	$\dfrac{1-2p}{[np(1-p)]^{1/2}}$	$3 + \dfrac{[1-6p(1-p)]}{np(1-p)}$

Para obtener una mejor perspectiva de la distribución binomial y de su forma, se calcularán α_3 y α_4 para distintos valores del parámetro n, de acuerdo con la tabla 4.2. Puede concluirse a partir de ésta, que la distribución binomial es simétrica si $p = 1/2$, con sesgo positivo si $p < 1/2$, y sesgada negativamente si $p > 1/2$. Para los últimos dos casos, el sesgo se vuelve menos evidente conforme n es más grande. Además, la distribución binomial es relativamente plana si $p = 1/2$. Para cualquier otro valor de p, la distribución binomial presenta un pico relativamente grande. Sin embargo, si n es grande α_4 tiende a tres para cualquier valor de p y la distribución es mesocúrtica.

De acuerdo con la definición 3.14, la función generadora de momentos para la distribución binomial es:

$$
\begin{aligned}
m_X(t) = E(e^{tX}) &= \sum_{x=0}^{n} e^{tx} \frac{n!}{(n-x)!x!} p^x(1-p)^{n-x} \\
&= \sum_{x=0}^{n} \frac{n!}{(n-x)!x!} (e^t p)^x(1-p)^{n-x} \\
&= (1-p)^n + n(1-p)^{n-1}(e^t p) \\
&\quad + \frac{n(n-1)}{2!}(1-p)^{n-2}(e^t p)^2 + \cdots + (e^t p)^n \\
&= [(1-p) + e^t p]^n. \qquad (4.15)
\end{aligned}
$$

TABLA 4.2 Factores de forma de la distribución binomial para distintos valores de p

	$p = 1/10$	$p = 1/2$	$p = 9/10$
α_3	$\dfrac{8}{3\sqrt{n}}$	0	$-\dfrac{8}{3\sqrt{n}}$
α_4	$3 + \dfrac{46}{9n}$	$3 - \dfrac{2}{n}$	$3 + \dfrac{46}{9n}$

Al tomar las dos primeras derivadas de (4.15) con respecto a t, se tiene:

$$\frac{dm_X(t)}{dt} = ne^t p[(1 - p) + e^t p]^{n-1}$$

y

$$\frac{d^2m_X(t)}{dt^2} = n(n - 1)(e^t p)^2[(1 - p) + e^t p]^{n-2} + ne^t p[(1 - p) + e^t p]^{n-1}.$$

Si $t = 0$, se obtienen los momentos primero y segundo alrededor del cero,

$$\left.\frac{dm_X(t)}{dt}\right|_{t=0} = np[(1 - p) + p]^{n-1}$$

$$= np$$

y:

$$\left.\frac{d^2m_X(t)}{dt^2}\right|_{t=0} = n(n - 1)p^2[(1 - p) + p]^{n-2} + np[(1 - p) + p]^{n-1}$$

$$= n(n - 1)p^2 + np,$$

que son idénticos a los determinados mediante el empleo del método directo. Los momentos de orden superior pueden determinarse mediante la continuación de este proceso de diferenciación y al evaluar la derivada en $t = 0$. Nótese que para este caso los primeros dos momentos alrededor del cero se obtienen de manera más fácil empleando la función generadora de momentos que tiene el método directo. Sin embargo, esto no ocurre en general.

Ejemplo 4.3 Un club nacional de automovilistas comienza una campaña telefónica con el propósito de aumentar el número de miembros. Con base en experiencia previa, se sabe que una de cada 20 personas que reciben la llamada se une al club. Si en un día 25 personas reciben la llamada telefónica ¿cuál es la probabilidad de que por lo menos dos de ellas se inscriban al club? ¿Cuál es el número esperado?

Puesto que una de cada 20 personas se suscriben al club, $p = 0.05$. Además, si se supone que las 25 personas constituyen un conjunto de ensayos independientes (una suposición muy razonable en este caso) con una probabilidad constante $p = 0.05$ de suscribirse al club, y si la variable aleatoria X es el número, de entre $n = 25$, que termina suscribiéndose al club, la probabilidad deseada es:

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - F(1; 25, 0.05) = 0.3576.$$

Mediante el empleo de (4.4), el valor esperado de X es $E(X) = (25)(0.05) = 1.25$.

4.3 La distribución de Poisson

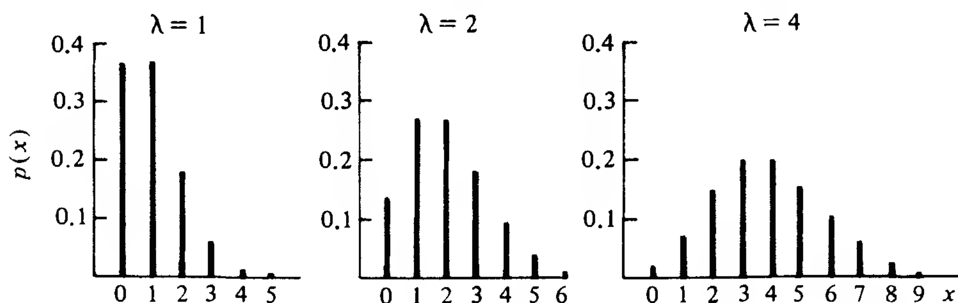
Llamada así en honor de Siméon Denis Poisson, probabilista francés del siglo XIX. quien fue el primero en describirla, es otra distribución discreta de probabilidad muy útil en la que la variable aleatoria representa el número de eventos independientes que ocurren a una velocidad constante. Muchos eventos aleatorios ocurren de manera independiente con una velocidad constante en el tiempo o en el espacio. Algunos ejemplos típicos son el número de personas que llegan a una tienda de autoservicio en un tiempo determinado, el número de defectos en piezas similares para el material, el número de bacterias en un cultivo, el número de solicitudes de seguro procesadas por una compañía en un periodo específico, etc. De hecho, la distribución de Poisson es el principal modelo de probabilidad empleado para analizar problemas de líneas de espera. Además, ofrece una aproximación excelente a la función de probabilidad binomial cuando p es pequeño y n grande. La deducción de la función de probabilidad de Poisson se desarrolla en un apéndice que se encuentra al final de este capítulo.

Definición 4.2 Sea X una variable aleatoria que representa el número de eventos aleatorios independientes que ocurren a una rapidez constante sobre el tiempo o el espacio. Se dice entonces que la variable aleatoria X tiene una distribución de Poisson con función de probabilidad.

$$p(x; \lambda) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} & x = 0, 1, 2, \ldots; \quad \lambda > 0, \\ 0 & \text{para cualquier otro valor.} \end{cases} \tag{4.16}$$

El parámetro de la distribución de Poisson es λ, el número promedio de ocurrencias del evento aleatorio por unidad de tiempo. Para valores mayores que cero, λ define una familia de distribuciones con una función de probabilidad determinada por (4.16). En la figura 4.2 se proporcionan algunas gráficas de la función de probabilidad de Poisson, para distintos valores de λ:

FIGURA 4.2 Gráficas de la función de probabilidad de Poisson



Puede verificarse que (4.16) es una función de probabilidad, puesto que $p(x; \lambda) > 0$ para $x = 0, 1, 2...$ y

$$\begin{aligned}\sum_{x=0}^{\infty} p(x; \lambda) &= \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \\ &= e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2!} + \cdots\right) \\ &= e^{-\lambda}e^{\lambda} \\ &= 1.\end{aligned}$$

Para ilustrar, sea $\lambda = 1.2$; entonces

$$p(x; 1.2) = \frac{e^{-1.2}1.2^x}{x!}, \quad x = 0, 1, 2, \ldots .$$

De esta forma se tiene

$$p(0; 1.2) = \frac{e^{-1.2}1.2^0}{0!} = 0.3012, \qquad p(4; 1.2) = \frac{e^{-1.2}1.2^4}{4!} = 0.0260,$$

$$p(1; 1.2) = \frac{e^{-1.2}1.2^1}{1!} = 0.3614, \qquad p(5; 1.2) = \frac{e^{-1.2}1.2^5}{5!} = 0.0062,$$

$$p(2; 1.2) = \frac{e^{-1.2}1.2^2}{2!} = 0.2169, \qquad p(6; 1.2) = \frac{e^{-1.2}1.2^6}{6!} = 0.0012,$$

$$p(3; 1.2) = \frac{e^{-1.2}1.2^3}{3!} = 0.0867, \qquad p(7; 1.2) = \frac{e^{-1.2}1.2^7}{7!} = 0.0002.$$

A pesar de que puede continuarse este proceso sin finalizar, nótese que las probabilidades individuales son más y más pequeñas conforme la variable aleatoria toma valores cada vez más grandes. Ésta es una característica general de la distribución de Poisson.

La probabilidad de que una variable aleatoria de Poisson X sea menor o igual a un valor de x se determina por la función de distribución acumulativa.

$$P(X \leq x) = F(x; \lambda) = \sum_{i=0}^{x} \frac{e^{-\lambda}\lambda^i}{i!}. \tag{4.17}$$

En la tabla B del apéndice, se encuentra tabulada (4.17) para distintos valores de x y λ. Nótese de nuevo que la variable aleatoria de Poisson tiene un valor entero, y que pueden usarse los valores de las probabilidades acumulativas de la tabla B para de-

terminar las probabilidades individuales mediante el empleo de la relación:

$$p(x; \lambda) = F(x; \lambda) - F(x - 1; \lambda).$$

A continuación se dan varios ejemplos del empleo de la tabla B. Sea $\lambda = 2.5$. La probabilidad de que X sea menor que tres es:

$$P(X < 3) = P(X \leq 2) = F(2; 2.5) = 0.5438;$$

la probabilidad de que X sea mayor que cuatro es:

$$P(X \geq 4) = 1 - P(X \leq 3) = 1 - F(3; 2.5) = 0.2424;$$

y la probabilidad de que X tome el valor de dos está dada por:

$$p(2; 2.5) = F(2; 2.5) - F(1; 2.5) = 0.2565.$$

Ejemplo 4.4 Después de una prueba de laboratorio muy rigurosa con cierto componente eléctrico, el fabricante determina que en promedio, sólo fallarán dos componentes antes de tener 1 000 horas de operación. Un comprador observa que son cinco los que fallan antes de las 1 000 horas. Si el número de componentes que fallan es una variable aleatoria de Poisson, ¿existe suficiente evidencia para dudar de la conclusión del fabricante?

La duda en estadística puede apoyarse en términos de la probabilidad. Si un evento debe o no ocurrir bajo ciertas condiciones, su ocurrencia se decide en términos de la probabilidad del evento bajo esas condiciones. Si la probabilidad de ocurrencia es pequeña y el evento ocurre, entonces se puede preguntar, con justificación, por las condiciones. Al mismo tiempo debe tenerse en mente que un valor de probabilidad pequeño no impide la ocurrencia del evento, a menos que este valor sea cero. En dicho caso, se tiene que $\lambda = 2$. Se supone que la frecuencia con que ocurren las fallas es constante e igual a dos por cada mil horas o un promedio de 1/500 unidades por hora. La probabilidad de que fallen cinco componentes en mil horas es:

$$p(5; 2) = \frac{e^{-2}2^5}{5!} = 0.0361,$$

y la probabilidad de que por lo menos fallen cinco en 1 000 horas es:

$$P(X \geq 5) = 1 - F(4; 2) = 0.0527.$$

Ambas probabilidades son, de manera relativa, pequeñas. Esto es, si el número de fallas en mil horas está descrita de manera apropiada por la distribución de Poisson con una frecuencia constante de dos, existe una probabilidad de observar exactamente cinco unidades defectuosas de 0.0361 y una probabilidad de 0.0527 de observar por lo menos cinco en el mismo periodo de operación. Sin embargo, antes de tomar cualquier medida en contra del fabricante, es necesario contestar algunas pre-

guntas. Por ejemplo, ¿es la frecuencia de falla constante e igual a dos durante mil horas? Aun si lo anterior fuese cierto, ¿es el medio de operación el mismo bajo el cual el fabricante hizo sus pruebas? Esto es, ¿es posible tener factores extraños, introducidos de manera inadvertida, que estén causando un número tan alto de fallas? Las preguntas anteriores sólo pueden constestarse con una comprensión completa de la situación.

Ejemplo 4.5 Considérese el juego de fútbol que se efectúa entre los 28 equipos que constituyen la Liga Nacional de Fútbol (NFL). Sea la variable aleatoria de interés el número de anotaciones — seis puntos (touchdowns) — de cada equipo por juego. Con el presente número de anotaciones por equipo en la temporada de 1979, ¿existe alguna razón para creer que el número de anotaciones es una variable aleatoria de Poisson?

Para contestar a esta pregunta, se compararán los resultados observados con los que se esperarían si el número de anotaciones fuese una variable aleatoria de Poisson, como se muestra en la tabla 4.3. La cuarta columna indica la probabilidad teórica para cada uno de los valores que aparecen en la primera columna, suponiendo que el número de anotaciones es una variable aleatoria de Poisson.

Los valores de la cuarta columna se determinan con el cálculo del valor del parámetro λ de la distribución de Poisson y la evaluación de la función de probabilidad (4.16) para los valores de la columna uno. El valor de λ se obtiene sumando los productos de las correspondientes posiciones de la primera y tercera columnas,

$$\begin{aligned}\lambda &= (0)(0.0781) + (1)(0.2210) + \cdots + (7)(0.0067)\\ &= 2.435\end{aligned}$$

TABLA 4.3 Distribución del número de anotaciones de seis puntos por equipo y por juego en la NFL, durante la temporada de 1979

Número de anotaciones	*Número de veces observadas*	*Frecuencia relativa*	*Probabilidad teórica*	*Número esperado de ocurrencias*
0	35	0.0781	0.0876	39.24
1	99	0.2210	0.2133	95.56
2	104	0.2321	0.2597	116.34
3	110	0.2455	0.2108	94.44
4	62	0.1384	0.1283	57.48
5	25	0.0558	0.0625	28.00
6	10	0.0223	0.0254	11.38
7*	3	0.0067	0.0124	5.56
Totales	448	0.9999	1.0000	448

* En realidad, esta cifra representa siete o más anotaciones, pero su ocurrencia es definitivamente escasa en la NFL.

lo que representa el números promedio de anotaciones por equipo y por juego. Las probabilidades puntuales se calculan mediante el empleo de:

$$p(x; 2.435) = \frac{e^{-2.435}(2.435)^x}{x!} \qquad x = 0, 1, 2, \ldots .$$

Éstos son los primeros siete renglones de la cuarta columna. El último renglón es la probabilidad de que X sea mayor o igual a siete. Los renglones de la última columna se encuentran multiplicando cada renglón de la columna cuatro por 448.

La comparación de las columnas dos y cinco, o de las columnas tres y cuatro, revela una concordancia muy razonable. Por lo tanto, puede concluirse que el número de anotaciones es una variable aleatoria de Poisson. Que la variable aleatoria sea del tipo Poisson, se basa en que el número de anotaciones por equipo y por juego en la NFL es un conjunto de eventos aleatorios independientes, de manera que la frecuencia de anotación es constante durante los 60 minutos del juego. La frecuencia de anotación puede ser más constante en la NFL como consecuencia de la calidad del juego y del oponente que en el fútbol colegial.

La distribución de Poisson también es una forma límite de la distribución binomial cuando $n \to \infty$ y $p \to 0$ de manera que no permanece constante. Este resultado se obtiene mediante el siguiente teorema, formulado por Siméon Poisson.

Teorema 4.1 Sea X una variable aleatoria con distribución binomial y función de probabilidad:

$$p(x; n, p) = \frac{n!}{(n-x)!x!} p^x(1-p)^{n-x} \qquad x = 0, 1, 2, \ldots n.$$

Si para $n = 1, 2 \ldots$ la relación $p = \lambda/n$ es cierta para alguna constante $\lambda > 0$, entonces:

$$\lim_{\substack{n \to \infty \\ p \to 0}} p(x; n, p) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots .$$

La prueba del teorema 4.1 se proporciona en un apéndice al final del capítulo.

En el contexto del teorema 4.1, la distribución de Poisson se piensa como aquélla en la que la variable aleatoria puede tomar valores distintos (n es grande pero las probabilidades son pequeñas) y — $p = \lambda/n$ tiene un valor cercano a cero. Como resultado, la función de probabilidad de Poisson se emplea de manera extensa para aproximar la función de probabilidad binomial cuando n es, de manera relativa, grande y p pequeño, de manera tal que $\lambda = np$ tiene un valor moderado. En la tabla 4.4. se ilustra la mejoría en la aproximación Poisson de la función de probabilidad binomial conforme n crece y p decrece tal que $\lambda = np$ permanece constante e igual a dos.

Ejemplo 4.6 Un comprador de grandes cantidades de circuitos integrados ha adoptado un plan para aceptar un envío de éstos y que consiste en inspeccionar una

TABLA 4.4 Comparación de las probabilidades binomial y de Poisson

	Binomial				de Poisson
x	$p(x;\ 10,\ 0.2)$	$p(x;\ 20,\ 0.1)$	$p(x;\ 40,\ 0.05)$	$p(x;\ 100,\ 0.02)$	$p(x;\ 2)$
0	0.1074	0.1216	0.1285	0.1326	0.1353
1	0.2684	0.2702	0.2706	0.2707	0.2707
2	0.3020	0.2852	0.2777	0.2734	0.2707
3	0.2013	0.1901	0.1851	0.1823	0.1804
4	0.0881	0.0898	0.0901	0.0902	0.0902
5	0.0264	0.0319	0.0342	0.0353	0.0361
6	0.0055	0.0089	0.0105	0.0114	0.0120
7	0.0008	0.0020	0.0027	0.0031	0.0034
8	0.0001	0.0004	0.0006	0.0007	0.0009
9	0.0000	0.0001	0.0001	0.0002	0.0002

muestra aleatoria de 100 circuitos provenientes del lote. Si el comprador encuentra no más de dos circuitos defectuosos en la muestra, acepta el lote; de otra forma, lo rechaza. Si se envía al comprador un lote que contiene 1% de circuitos defectuosos, ¿cuál es la probabilidad de que éste sea aceptado?

Sea X la variable aleatoria que representa el número de circuitos defectuosos encontrados en una muestra de 100 y supóngase que X tiene una distribución binomial. En otras palabras, se supone que los 100 circuitos seleccionados del lote constituyen 100 ensayos independientes, de manera tal que la probabilidad de tener un circuito defectuoso es constante e igual a 0.01. La probabilidad de aceptar el lote es la misma de X con valor menor o igual a dos. Dado que $n = 100$ es relativamente un valor grande y $p = 0.01$ es pequeño; la probabilidad binomial puede aproximarse mediante la distribución de Poisson, escogiendo $\lambda = np = 1$:

$$P(\text{aceptación}) \equiv P(X \leq 2) = F_P{}^*(2;\ 1) = 0.9197.$$

Debe notarse por comparación que si se empleara la distribución binomial se tendría:

$$P(X \leq 2) = F_B{}^*(2;\ 100,\ 0.01) = 0.9206.$$

Los momentos de la variable aleatoria de Poisson se determinan mediante los mismos procedimientos utilizados para obtener los momentos de la variable aleatoria binomial. Si X es una variable aleatoria de Poisson, su valor esperado es:

$$E(X) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!}$$

* Se emplean los subíndices para distinguir entre las dos funciones de distribución. Se emplearán las mismas marcas para distinguir entre dos funciones de probabilidad, cuando sea necesario.

$$= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

$$= \lambda e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^{y}}{y!}, \qquad y = x - 1$$

$$= \lambda. \tag{4.18}$$

Para la varianza X:

$$E[X(X-1)] = \sum_{x=0}^{\infty} x(x-1) \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$= \lambda^2 e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!}$$

$$= \lambda^2 e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^{y}}{y!}, \qquad y = x - 2$$

$$= \lambda^2. \tag{4.19}$$

Entonces, de (4.5):

$$E(X^2) = \mu_2' = \lambda^2 + \lambda,$$

y la varianza de X es:

$$Var(X) = \mu_2' - \mu^2$$

$$= \lambda^2 + \lambda - \lambda^2$$

$$= \lambda. \tag{4.20}$$

De esta manera, una característica distintiva de la variable aleatoria de Poisson es que su media es igual a su varianza.

El ejercicio para el lector es que demuestre que, para el tercer momento central, se tiene:

$$E[X(X-1)(X-2)] = \lambda^3. \tag{4.21}$$

Mediante el empleo de (4.7):

$$\mu_3' = \lambda^3 + 3\lambda^2 + \lambda,$$

y el tercer momento central es:

$$\mu_3 = \lambda.$$

Como resultado, el coeficiente de asimetría se determina por:

$$\alpha_3 = \mu_3/\mu_2^{3/2} = 1/\sqrt{\lambda}. \tag{4.22}$$

Para el cuarto momento central puede emplearse el mismo procedimiento para demostrar que:

$$E[X(X-1)(X-2)(X-3)] = \lambda^4, \tag{4.23}$$

y de (4.8):

$$\mu_4' = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda. \tag{4.24}$$

Mediante el empleo de (3.10) el cuarto momento central es:

$$\mu_4 = 3\lambda^2 + \lambda,$$

y el cuarto momento estandarizado para la distribución de Poisson lo establece:

$$\alpha_4 = \mu_4/\mu_2^2 = 3 + \frac{1}{\lambda}. \tag{4.25}$$

Se proporciona un resumen de las propiedades de la distribución de Poisson en la tabla 4.5. La distribución de Poisson se encuentra sesgada positivamente para cualquier valor $\lambda > 0$, pero la asimetría disminuye para valores relativamente grandes de λ. Además, la distribución de Poisson es leptocúrtica, puesto que α_4 es mayor que tres, pero tiende a convertirse en mesocúrtica para valores grandes de λ.

La función generadora de momentos para la distribución de Poisson se determina por:

$$\begin{aligned} m_X(t) &= \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} \\ &= e^{-\lambda} e^{\lambda e^t} \\ &= \exp[\lambda(e^t - 1)]. \end{aligned} \tag{4.26}$$

TABLA 4.5 Propiedades básicas de la distribución de Poisson

Función de probabilidad	*Parámetro*
$p(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$	$\lambda > 0$

Media	*Varianza*	*Coeficiente de asimetría*	*Curtosis relativa*
λ	λ	$\frac{1}{\sqrt{\lambda}}$	$3 + \frac{1}{\lambda}$

Nótese que, como se esperaba: $m_X(0) = e^{\lambda(1-1)} = 1$. El ejercicio para el lector es demostrar que (4.26) da los momentos de la variable aleatoria de Poisson después de llevar a cabo el proceso de diferenciación apropiado.

En conclusión, la distribución de Poisson es leptocúrtica con un sesgo positivo y se emplea para modelar el número de eventos aleatorios independientes que ocurren a una rapidez constante ya sea sobre el tiempo o el espacio. Se ha empleado de manera extensa para el estudio de línea de espera, confiabilidad y control de calidad. Es también una forma límite de la distribución binomial y la aproxima de manera adecuada para valores grandes de n y pequeños de p. Sin embargo, debe aplicarse cuidadosamente la distribución de Poisson a situaciones en las que las condiciones de independencia y rapidez constante de ocurrencia son dudosas.

Por ejemplo, considérese la distribución del número de infracciones recibidas por los automovilistas en un periodo de diez años. Puede argumentarse que la distribución de Poisson es el modelo de probabilidad adecuado, pues la probabilidd de recibir una infracción en un día cualquiera es pequeña y hay muchos días en diez años. Sin embargo, no es común que las condiciones de independencia y rapidez constante sean válidas. La independencia es dudosa debido a que si un automovilista en particular recibe una infracción, es razonable pensar que manejará de manera más cuidadosa. En grupos de distinta edad esta frecuencia puede variar, ya que las compañías aseguradoras sostienen que los conductores de mayor edad respetan más los límites de velocidad que los conductores jóvenes.

4.4. La distribución hipergeométrica

Para establecer las condiciones básicas que llevan a otra distribución discreta de probabilidad conocida como hipergeométrica, considérese el siguiente problema: sea N el número de representantes de un determinado estado que asisten a una convención política nacional, y sea k el número de los que apoyan al candidato A, mientras que el resto $N-k$ apoya al candidato B. Supóngase que una organización informativa selecciona aleatoriamente a n representantes y les pregunta sus razones para apoyar a los candidatos. Si X es una variable aleatoria que sustituye el número de representantes en la muestra que apoyan al candidato A, ¿cuál es la función de probabilidad de X?

Esta situación parece ser binomial porque entre N representantes de un estado existen dos grupos distintos con probabilidad k/N y $(N-k)/N$. Sin embargo, considérese con más detalle el proceso de selección para la muestra de n representantes. Es razonable suponer que se selecciona un representante, se le preguntan sus razones y no vuelve a ser seleccionado.* El resultado es que no existe independencia entre la selección de un representante y el siguiente. Por ejemplo, supóngase que el primer representante seleccionado apoya al candidato A. Entonces quedan $N-1$ representantes de los cuales $k-1$ apoya a A. Por lo tanto, la probabilidad condicional de que

* Esto se conoce como *muestreo sin reemplazo* y es una condición fundamental para la distribución hipergeométrica. En la distribución binomial, se supone que el muestreo se hace con reemplazo, asegurando la independencia y la probabilidad constante.

el siguiente candidato apoye también a A es $(k-1)/(N-1)$ y no k/N, y la probabilidad condicional de que el siguiente representante apoye a B es $(N-k)/(N-1)$ y no $(N-k)/N$.

Para determinar la probabilidad de que, de maneras exacta, se seleccionen x representantes que apoyen a A y $n-x$ que apoyen a B, se procederá de la siguiente forma: el número de maneras distintas en que puede seleccionarse una muestra de n representantes de un total de N es $\binom{N}{n}$; y cada muestra tiene una probabilidad de selección igual a $1/\binom{N}{n}$. De manera similar, la selección de x personas que apoyen a A es un evento que puede ocurrir de $\binom{k}{x}$ maneras distintas, y la selección de $(n-x)$ representantes que apoyen a B es un evento que puede suceder de $\binom{N-k}{n-x}$ maneras. El número total de maneras en que ambos eventos pueden ocurrir es $\binom{k}{x}\binom{N-k}{n-x}$. De esta forma, la probabilidad de seleccionar x representantes que apoyen al candidato A es

$$p(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}.$$

Definición 4.3 Sea N el número total de objetos en una población finita, de manera tal que k de éstos es de un tipo y $N-k$ de otros. Si se selecciona una muestra aleatoria* de la población constituida por n objetos de la probabilidad de que x sea de un tipo exactamente y $n-x$ sea del otro, está dada por la función de probabilidad hipergeométrica:

$$p(x; N, n, k) = \begin{cases} \dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, & x = 0, 1, 2, \ldots, n; \quad x \leq k, \quad n - x \leq N - k; \\ & N, n, k, \text{ enteros positivos,} \\ 0 & \text{para cualquier otro valor} \end{cases} \tag{4.27}$$

Los parámetros de la distribución hipergeométrica son N, n, y k. Éstos definen una familia de distribuciones con función de probabilidad determinada por (4.27). En la figura 4.3 se muestran algunas gráficas de (4.27) para distintas combinaciones de N, n, y k.

La función de probabilidad (4.27) de la distribución hipergeométrica y la función de distribución acumulativa, definida por:

$$P(X \leq x) = F(x; N, n, k) = \sum_{i=0}^{x} \frac{\binom{k}{i}\binom{N-k}{n-i}}{\binom{N}{n}}, \tag{4.28}$$

* Véase el capítulo siete para la definición de una muestra aleatoria.

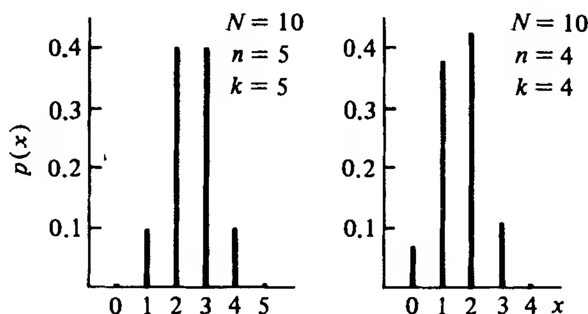


FIGURA 4.3 Gráficas de la función hipergeométrica de probabilidad

se encuentra tabulada en [4] para valores de N, n, y k desde $N = 2$, $n = 1$ hasta $N = 100$ y $n = 50$. Una parte de éstas se encuentra en la tabla C del apéndice. El cálculo de las probabilidades hipergeométricas puede convertirse en tedioso, especialmente si n es grande. Sin embargo, puede simplificarse si se emplea la siguiente fórmula de recursión,

$$p(x+1; N, n, k) = \frac{(n-x)(k-x)}{(x+1)(N-k-n+x+1)} p(x; N, n, k), \qquad (4.29)$$

la cual se puede obtener directamente de la función de probabilidad hipergeométrica.

Ejemplo 4.7 Supóngase que se tienen 50 representantes de cierto estado, a una convención política nacional, de los cuales 30 apoyan al candidato A y 20 al candidato B. Si se seleccionan aleatoriamente cinco representantes, ¿cuál es la probabilidad de que, entre estos cinco, por lo menos dos apoyen al candidato A?

Sea X la variable aleatoria que representa el número de personas en la muestra que apoyan a A. Para $N = 50$, $n = 5$, y $k = 30$, la función de probabilidad de X está dada por:

$$p(x; 50, 5, 30) = \frac{\binom{30}{x}\binom{20}{5-x}}{\binom{50}{5}}, \quad x = 0, 1, \dots, 5,$$

y la probabilidad de que $X \geqslant 2$ es:

$$P(X \geqslant 2) = 1 - P(X \leqslant 1) = 1 - [p(0; 50, 5, 30) + p(1; 50, 5, 30)].$$

Dado que:

$$p(0; 50, 5, 30) = \frac{\binom{30}{0}\binom{20}{5}}{\binom{50}{5}} = \frac{\binom{20}{5}}{\binom{50}{5}} = 0.007317,$$

y, de (4.29):

$$p(1; 50, 5, 30) = \frac{(5 - 0)(30 - 0)}{(0 + 1)(50 - 30 - 5 + 0 + 1)} p(0; 50, 5, 30) = 0.068597,$$

se encuentra que:

$$P(X \geq 2) = 1 - (0.007317 + 0.068597) = 0.9241.$$

Un área muy fructífera en aplicaciones para la distribución hipergeométrica es el control estadístico de calidad y la aceptación de muestreo. En este contexto sea N el número de unidades en un lote, de las cuales k se encuentran defectuosas. Si se selecciona una muestra aleatoria del lote formada por $n < N$ unidades, la probabilidad de que la muestra contenga x unidades defectuosas se determina mediante el empleo de la función hipergeométrica de probabilidad (4.27). En aceptación del muestreo, la razón de que sólo se seleccione la muestra de un lote obedece más bien a restricciones de tiempo y dinero. La decisión de cuándo aceptar o rechazar un lote se basa, de manera general, en el número de artículos defectuosos encontrados en él. Estos conceptos se tratarán con gran detalle en el capítulo once.

Ejemplo 4.8 Considérese un fabricante de automóviles que compra los motores a una compañía donde se fabrican bajo estrictas especificaciones. El fabricante recibe un lote de 40 motores. Su plan para aceptar el lote consiste en seleccionar ocho, de manera aleatoria, y someterlos a prueba. Si encuentra que ninguno de los motores presenta serios defectos, el fabricante acepta el lote; de otra forma lo rechaza. Si el lote contiene dos motores con serios defectos, ¿cuál es la probabilidad de que sea aceptado?

Sea X el número de motores defectuosos en la muestra. Para $N = 40$, $n = 8$, y $k = 2$, la probabilidad de aceptación es

$$p(0; 40, 8, 2) = \frac{\binom{2}{0}\binom{38}{8}}{\binom{40}{8}} = 0.6359.$$

De esta manera el lote 40 tiene una probabilidad menor de 2/3 de ser aceptado si contiene dos motores defectuosos. Debe notarse que la esencia del control estadístico de calidad es la mejoría de la calidad del producto. Si un vendedor sabe

que su producto pasará por una selección que verifica la calidad del producto, puede poner en marcha en su propia fábrica un control de calidad intencionado con el propósito de minimizar el número de lotes rechazados. Por lo tanto, es razonable suponer que esta práctica dará como resultado un producto de calidad superior.

¿Qué pasa con la distribución hipergeométrica si el tamaño de la muestra n es sólo una pequeña fracción de un lote de tamaño N relativamente grande? Supóngase que se envía un lote de 2 mil unidades de las cuales 40 se encuentran defectuosas. Si se selecciona una muestra de 50, sin reemplazo, la probabilidad de que el primer artículo seleccionado se encuentre defectuoso es de 40/2 000 = 0.02. La probabilidad condicional de que el segundo artículo también se encuentre defectuoso dado que el primero lo fue, es 39/1 999 = 0.0195. A pesar de que estas probabilidades no tienen el mismo valor, puede argumentarse, desde un punto de vista práctico, que la diferencia es insignificante. Es por esta razón que en muchas ocasiones se emplea la distribución binomial para aproximar a la distribución hipergeométrica cuando el cociente n/N es pequeño.

Si la proporción de artículos defectuosos en el lote es $p = k/N$, puede escribirse la función de probabilidad hipergeométrica como:

$$p_H(x; N, n, p) = \frac{\binom{Np}{x}\binom{N - Np}{n - x}}{\binom{N}{n}} \tag{4.30}$$

Puede demostrarse entonces que

$$\lim_{N \to \infty} p_H(x; N, n, p) = p_B(x; n, p),$$

en donde $p_B(x; n, p)$ es la función de probabilidad binomial. De esta forma la distribución hipergeométrica tiende a la binomial con parámetros n y $p/k/N$ conforme el cociente n/N se vuelve más pequeño. De manera general, la función de probabilidad binomial aproximará de manera adecuada a (4.30) si se tiene que $n < 0.1N$. En la tabla 4.6 se proporcionan algunas comparaciones entre las probabilidades binomial e hipergeométrica conforme el cociente n/N disminuye.

Ejemplo 4.9 Un fabricante asegura que sólo el 1% de su producción total se encuentra defectuosa. Supóngase que se ordenan 100 artículos y se seleccionan 25 al azar para inspeccionarlos. Si el fabricante se encuentra en lo correcto, ¿cuál es la probabilidad de observar dos o más artículos defectuosos en la muestra?

Sea X el número de artículos defectuosos en la muestra. Entonces X es una variable aleatoria hipergeométrica con parámetros N = 1 000, n = 25, y $k = Np$ = (1 000)(0.01) = 10. Dado que el cociente n/N es, de forma considerable, menor de 0.1, puede emplearse la distribución binomial para aproximar la probabilidad deseada:

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - F_B(1; 25, 0.01) = 0.0258.$$

TABLA 4.6 Comparación entre los valores de probabilidad binomial o hipergeométrica

x	Hipergeométrica $p(x;\ 100,\ 20,\ 5)$	Binomial $p(x;\ 20,\ 0.05)$	Hipergeométrica $p(x;\ 100,\ 10,\ 5)$	Binomial $p(x;\ 10,\ 0.05)$	Hipergeométrica $p(x;\ 100,\ 5,\ 5)$	Binomial $p(x;\ 5,\ 0.05)$
0	0.3193	0.3585	0.5838	0.5987	0.7696	0.7738
1	0.4201	0.3774	0.3394	0.3151	0.2114	0.2036
2	0.2073	0.1887	0.0702	0.0746	0.0184	0.0214
3	0.0478	0.0596	0.0064	0.0105	0.0006	0.0011
4	0.0051	0.0133	0.0003	0.0010	0.0000	0.0000
5	0.0002	0.0022	0.0000	0.0001	0.0000	0.0000

en donde $F_B(1;\ 25,\ 0.01)$ es la función de distribución acumulativa binomial. A continuación se analizará el proceso de decisión para este problema. La probabilidad de tener dos o más artículos defectuosos en la muestra es muy pequeña. Supóngase que se observan dos o más artículos defectuosos; entonces el proceso de decisión relativo al lote debe hacerse con base en la probabilidad. Esto es, si se supone que las condiciones son verdaderas, se ha observado algo que sólo tenía una oportunidad de 2.5% de ocurrir. Por otro lado, si la aseveración del fabricante no es cierta y la proporción de artículos defectuosos es del 3%, entonces la probabilidad de observar dos o más defectuosos es

$$P(X \geq 2) = 1 - F(1; 25, 0.03) = 0.1720,$$

que es un valor más plausible a la luz de la evidencia actual que es de 0.0258. De esta forma, si se observan dos o más artículos defectuosos de entre los 25, se debe rechazar el lote.

Para determinar la media de la distribución hipergeométrica se sigue un procedimiento análogo al empleado para la distribución binomial. Si la función de probabilidad está dada por (4.27),

$$E(X) = \sum_{x=0}^{n} x \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$= \sum_{x=1}^{n} x \frac{\frac{k!}{(k-x)!x!}\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$= k \sum_{x=1}^{n} \frac{\binom{k-1}{x-1}\binom{N-k}{n-x}}{\binom{N}{n}};$$

pero puede demostrarse que:

$$\binom{N}{n} = \frac{N}{n}\binom{N-1}{n-1},$$

o:

$$\frac{N!}{(N-n)!n!} = \frac{N}{n}\left[\frac{(N-1)!}{(N-n)!(n-1)!}\right].$$

Entonces:

$$E(X) = k\sum_{x=1}^{n}\frac{\binom{k-1}{x-1}\binom{N-k}{n-x}}{\frac{N}{n}\binom{N-1}{n-1}}$$

$$= \frac{nk}{N}\sum_{x=1}^{n}\frac{\binom{k-1}{x-1}\binom{N-k}{n-x}}{\binom{N-1}{n-1}}$$

Si $M = N - 1$, $r = k - 1$, $s = n - 1$ y $y = x - 1$,

$$E(X) = \frac{nk}{N}\sum_{y=0}^{s}\frac{\binom{r}{y}\binom{M-r}{s-y}}{\binom{M}{s}}$$

$$= \frac{nk}{N}; \qquad (4.31)$$

la suma es igual a uno dado que es la suma de una función de probabilidad hipergeométrica con parámetros M, s, y r. Nótese que si $p = k/N$, la media de la variable aleatoria hipergeométrica es la misma que la de la variable aleatoria binomial.

Con el mismo procedimiento puede demostrarse que la varianza de una distribución hipergeométrica es:

$$Var(X) = \frac{nk(N-k)}{N^2}\cdot\frac{(N-n)}{(N-1)}. \qquad (4.32)$$

Si $p = k/N$ y $(1 - p) = (N - k)/N$,

$$Var(X) = np(1-p)\left(\frac{N-n}{N-1}\right)$$

La varianza de una variable aleatoria hipergeométrica es más pequeña que la corres-

pondiente a la variable aleatoria binomial por un factor de $(N - n)/(N - 1)$. Sin embargo, si N es grande al compararse con n, este factor se encontrará cercano a uno, dando como resultado una varianza prácticamente igual a la binomial. El resultado anterior era de esperarse ya que si n es sólo una pequeña fracción de un lote de tamaño N, la distribución hipergeométrica tiende a la distribución binomial.

La determinación del coeficiente de asimetría y la curtosis relativa para la distribución hipergeométrica sigue el mismo procedimiento dado para la distribución binomial. Estas cantidades se dan en la tabla 4.7. Nótese que para $N > 2$, si $N < 2k$ o si $N < 2n$, la distribución hipergeométrica se encuentra sesgada negativamente. Si $N = 2k$ o si $N = 2n$, es simétrica. Si $N > 2k$ y $N > 2n$, la distribución se encuentra sesgada positivamente. El lector puede consultar [2] para la función generadora de momentos. Debe notarse que la función generadora de momentos representa un trabajo muy tedioso para determinar los momentos. La tabla 4.7 proporciona un resumen de la información más importante para esta distribución.

4.5 La distribución binomial negativa

Sea un escenario binomial en que se observa una secuencia de ensayos independientes; la probabilidad de éxito en cada ensayo es constante e igual a p. En lugar de fijar el número de ensayos en n y observar el número de éxitos, supóngase que se continúan los ensayos hasta que han ocurrido exactamente k éxitos. En este caso, la variable aleatoria es el número de ensayos necesarios para observar k éxitos. Esta situación lleva a lo que se conoce como la distribución binomial negativa.

TABLA 4.7 Propiedades básicas de la distribución hipergeométrica

Función de probabilidad	*Parámetros*
$p(x; N, n, k) = \dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$ $x = 0, 1, 2, \ldots, n$ $x \leq k, \quad n - x \leq N - k$	N, n, k, enteros positivos $1 \leq n \leq N; \quad 1 \leq k \leq N$ $N = 1, 2, \ldots$

Media	*Varianza*	*Coeficiente de asimetría*	*Curtosis relativa*
$\dfrac{nk}{N}$	$\dfrac{nk(N-k)(N-n)}{N^2(N-1)}$	$\dfrac{(N-2k)(N-2n)(N-1)^{1/2}}{(N-2)[nk(N-k)(N-n)]^{1/2}}$	*

$$* \; \alpha_4 = \frac{N^2(N-1)}{(N-2)(N-3)nk(N-k)(N-n)} \left\{N(N+1) - 6n(N-n) + 3\frac{k}{N^2}(N-k)\left[N^2(n-2) - Nn^2 + 6n(N-n)\right]\right\}$$

La determinación de la función de probabilidad sigue el mismo tipo de razonamiento empleado para obtener las funciones de probabilidad de las distribuciones binomial e hipergeométrica. Se desea determinar la probabilidad de que en el n-ésimo ensayo ocurra el k-ésimo éxito. Si se continúan los ensayos independientes hasta que ocurre el k-ésimo éxito, entonces el resultado del último ensayo fue éxito. Antes del último ensayo, habían ocurrido $k-1$ éxitos en $n-1$ ensayos. El número de maneras distintas en las que pueden observarse $k-1$ éxitos en $n-1$ ensayos es: $\binom{n-1}{k-1}$. Por lo tanto, la probabilidad de tener k éxitos en n ensayos con el último siendo un éxito, es:

$$p(n;\, k,\, p) = \binom{n-1}{k-1} p^k(1-p)^{n-k} \qquad n = k,\, k+1,\, k+2, \ldots . \qquad (4.33)$$

La expresión (4.33) es la función de probabilidad de lo que se conoce como la *distribución de Pascal.* Mediante el empleo de (4.33) puede obtenerse la distribución binomial negativa sustituyendo $n = x + k$ en (4.33), en donde x es el valor de una variable aleatoria que representa el número de fracasos hasta que se observan, de manera exacta, k éxitos.

Definición 4.4 Sea $X + k$, el número de ensayos independientes necesarios para alcanzar, de manera exacta, k éxitos en un experimento binomial en donde la probabilidad de éxito en cada ensayo es p. Se dice entonces que X es una variable binomial negativa con función de probabilidad

$$p(x;\, k,\, p) = \begin{cases} \binom{k+x-1}{k-1} p^k(1-p)^x & x = 0, 1, 2, \ldots \\ & k = 1, 2, \ldots \\ & 0 \leq p \leq 1, \\ 0 & \text{para cualquier otro valor} \end{cases} \qquad (4.34)$$

La distribución se llama "binomial negativa" debido a que las probabilidades dadas por (4.34) corresponden a los términos sucesivos de la expansión binomial de:

$$\left(\frac{1}{p} - \frac{1-p}{p}\right)^{-k}.$$

Los parámetros de la distribución binomial negativa son k y p, en donde k no necesita ser un entero. Si es así, la distribución se conoce como distribución de Pascal, misma que se interpreta como el tiempo que hay que esperar para que ocurra el k éxito. Si k no es entero, la función de probabilidad dada por (4.34) se escribe de manera tal que se involucre a la función gama,

$$p(x;\, k,\, p) = \frac{\Gamma(k+x)}{x!\,\Gamma(k)} p^k(1-p)^x \qquad \begin{matrix} x = 0, 1, 2, \ldots \\ k > 0, \quad 0 \leq p \leq 1. \end{matrix} \qquad (4.35)$$

En este contexto la distribución binomial negativa es un caso particular de la distri-

bución de Poisson compuesta. Una distribución compuesta de una variable aleatoria X es aquella que depende de un parámetro que a su vez es una variable aleatoria con una distribución dada. En el capítulo seis se plantea este problema para la distribución binomial negativa.

Debe notarse que si $k = 1$ en (4.34), surge un caso especial de la distribución binomial negativa, que se conoce con el nombre de *distribución geométrica* y cuya función de probabilidad está dada por

$$p(x; p) = p(1 - p)^x, \qquad x = 0, 1, 2, \ldots, \quad 0 \leq p \leq 1. \tag{4.36}$$

La variable aleatoria geométrica representa el número de fallas que ocurren antes de que se presente el primer éxito. En la figura 4.4 se ilustran varias gráficas de la función de probabilidad binomial negativa (4.34) para varios valores de k y p.

En la referencia [6] se encuentra una extensa tabla de probabilidades individual y acumulativas para la distribución binomial negativa. Es posible emplear la distribución binomial para obtener las probabilidades de la distribución binomial negativa. Puede demostrarse que si X es una variable aleatoria binomial negativa con función de probabilidad dada por (4.34), entonces:

$$P(X \leq x) = P(Y \geq k),$$

en donde Y es una variable aleatoria binomial con parámetros $n = k + x$ y p. Esto es:

$$F_{NB}(x; k, p) = 1 - F_B(k - 1; k + x, p), \tag{4.37}$$

en donde $F_{NB}(x; \ k, \ p)$ es la distribución binomial negativa acumulativa y $F_B(k - 1; k + x, p)$ es la distribución binomial acumulativa. Mediante el empleo de (4.37) puede determinarse las probabilidades individuales de la distribución binomial negativa. Por ejemplo,

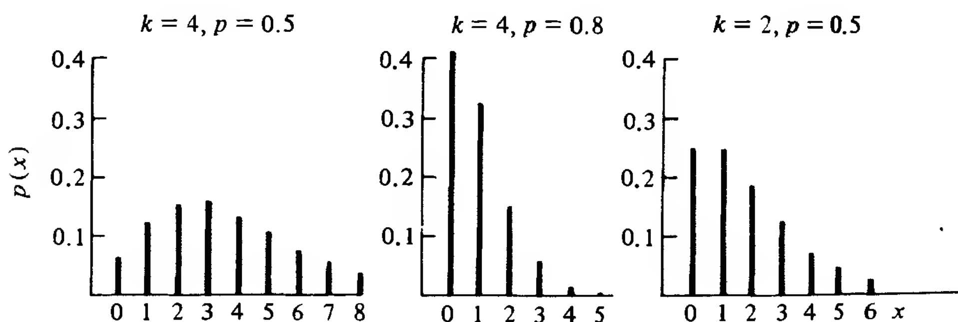


FIGURA 4.4 Gráficas de la función de probabilidad binomial negativa

$$\begin{aligned} P(X = x) &= F_{NB}(x; k, p) - F_{NB}(x - 1; k, p) \\ &= [1 - F_B(k - 1; k + x, p)] - [1 - F_B(k - 1; k + x - 1, p)] \\ &= F_B(k - 1; k + x - 1, p) - F_B(k - 1; k + x, p). \end{aligned} \quad (4.38)$$

Para ilustrar el uso de (4.37) y (4.34), sea $k = 2$ y $p = 0.5$ en (4.34):

$$p_{NB}(x; 2, 0.5) = (x + 1)(0.5)^2(0.5)^x, \quad x = 0, 1, 2, \ldots .$$

La probabilidad de que $X \leqslant 3$ es

$$P(X \leqslant 3) = F_{NB}(3; 2, 0.5) = 1 - F_B(1; 5, 0.5) = 0.8125;$$

la probabilidad de que $X = 2$ es

$$P(X = 2) = F_B(1; 3, 0.5) - F_B(1; 4, 0.5) = 0.1875;$$

y la probabilidad de que $X > 1$ es

$$\begin{aligned} P(X > 1) = P(X \geqslant 2) &= 1 - F_{NB}(1; 2, 0.5) \\ &= 1 - [1 - F_B(1; 3, 0.5)] \\ &= 0.5. \end{aligned}$$

La aplicación primaria de la distribución binomial negativa es una alternativa adecuada para el modelo de Poisson cuando la frecuencia de ocurrencia no es constante sobre el tiempo o el espacio. También se emplea de manera frecuente para modelar las estadísticas de accidentes, datos psicológicos, compras del consumidor y otras situaciones similares en donde la frecuencia de ocurrencia entre grupos o individuos no se espera que sea la misma. Por ejemplo, las estadísticas de accidentes automovilísticos indican de manera consistente que los conductores jóvenes tienen más accidentes que los de más edad, y que los hombres tienen un mayor número de accidentes que las mujeres. Desde este punto de vista no debe tomarse la distribución binomial negativa en términos de cuántos ensayos se necesitan para alcanzar un determinado número de éxitos. Más bien, debe considerarse como el número de ocurrencias en el tiempo o en el espacio cuando la frecuencia de éstas no es constante. Para una aplicación en particular, veáse la referencia [1].

Los momentos de una variable aleatoria binomial negativa pueden determinarse al obtener los momentos factoriales, como se hizo para las distribuciones binomial, de Poisson e hipergeométrica. También es posible obtener la media, la varianza, el coeficiente de asimetría y la curtosis relativa a partir de las expresiones dadas por (4.4), (4.6) y (4.14) respectivamente. Puede demostrarse que si estas expresiones reemplazan los parámetros binomiales n, $(1 - p)$ y p con las cantidades $-k$, $1/p$ y $-(1-p)/p$, respectivamente, se obtendrán los momentos binomiales negativos deseados. De acuerdo con lo anterior, si X es una variable aleatoria binomial negativa con función

de probabilidad dada por (4.34):

$$E(X) = \frac{k(1-p)}{p}, \tag{4.39}$$

$$Var(X) = \frac{k(1-p)}{p^2}, \tag{4.40}$$

$$\alpha_3 = \frac{2-p}{[k(1-p)]^{1/2}}, \qquad \text{y} \tag{4.41}$$

$$\alpha_4 = 3 + \frac{(p^2 - 6p + 6)}{k(1-p)}. \tag{4.42}$$

En la tabla 4.8 se proporciona la información más útil para la distribución binomial negativa. A partir de esta tabla son evidentes algunas propiedades básicas de tal distribución. La varianza es más grande que la media en forma permanente, así como la distribución presenta un sesgo positivo y es leptocúrtica puesto que α_4 siempre es más grande que tres, pero $\alpha_4 \to 3$ conforme $k \to \infty$.

Ejemplo 4.10 En un artículo de R. Pollard (véase la referencia [5]) se demuestra que el número de anotaciones de seis puntos por equipo en el fútbol colegial se describe de manera apropiada mediante una distribución binomial negativa. La tabla 4.9 contiene información muy semejante a la que aparece en la tabla 4.3. Para determinar de manera teórica la probabilidad de ocurrencia, es necesario tener estimaciones de los valores de los parámetros k y p. Dado que la media y la varianza de una variable aleatoria binomial negativa están dadas por (4.39) y (4.40) respectivamente, se resuelve para k y p y se obtiene:

$$p = \frac{E(X)}{Var(X)}, \qquad \text{y} \qquad k = \frac{E^2(X)}{Var(X) - E(X)}.$$

TABLA 4.8 Propiedades básicas de la distribución binomial negativa

Función de probabilidad	*Parámetros*
$p(x; k, p) = \binom{k+x-1}{k-1} p^k (1-p)^x$	k, $k > 0$ (distribución de Pascal si k es un entero positivo)
$x = 0, 1, 2, \ldots$	p, $0 \leq p \leq 1$

Media	*Varianza*	*Coeficiente de asimetría*	*Curtosis relativa*
$\frac{k(1-p)}{p}$	$\frac{k(1-p)}{p^2}$	$\frac{2-p}{[k(1-p)]^{1/2}}$	$3 + \frac{(p^2-6p+6)}{k(1-p)}$

TABLA 4.9 Distribución del número de anotaciones de seis puntos por equipo y por juego en el fútbol colegial, 1967

Número de anotaciones	*Número de veces observadas*	*Frecuencia relativa*	*Probabilidad teórica*	*Número esperado de ocurrencias*
0	272	0.1174	0.1205	279
1	485	0.2094	0.2117	490
2	537	0.2319	0.2197	509
3	407	0.1757	0.1754	406
4	258	0.1114	0.1190	276
5	157	0.0678	0.0722	167
6	101	0.0436	0.0404	94
7	57	0.0246	0.0212	49
8	23	0.0099	0.0106	25
9	8	0.0035	0.0051	12
10	5	0.0022	0.0023	5
11+	6	0.0026	0.0019	4
Totales	2316	1.0000	1.0000	2316

El método con que se calculan estos parámetros* es la suposición de que las estimaciones de $E(X)$ y Var(X) son iguales a la media $\bar{x}$ y la varianza s^2, muestral, mismas que tienen un valor de 2.58 y 3.79 respectivamente. De acuerdo con lo anterior, la estimación de p resulta ser 0.6807 y la de k, 5.5012. Puesto que esta última no es un entero, se emplea la función de probabilidad dada por (4.35) para determinar las probabilidades teóricas.

La diferencia aparente entre las distribuciones del número de anotaciones por equipo entre la NFL y el fútbol colegial se puede explicar en gran parte por la gran variabilidad que existe en la calidad de los oponentes en el fútbol colegial cuando éste se compara con la NFL. Como resultado, se espera que la frecuencia con la que se anotan seis puntos en el fútbol colegial sea más una función del oponente de lo que es en la NFL. De esta manera es como se sugiere la distribución binomial negativa.

Mediante un empleo directo de la definición, la función generadora de momentos de la distribución binomial negativa se obtiene de la siguiente manera:

$$E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} \binom{k+x-1}{k-1} p^k (1-p)^x$$

$$= \sum_{x=0}^{\infty} \frac{(k+x-1)!}{(k-1)!x!} p^k[(1-p)e^t]^x$$

$$= p^k + kp^k[(1-p)e^t] + \frac{k(k+1)}{2!} p^k[(1-p)e^t]^2 + \cdots,$$

* Véase el capítulo ocho, en particular la sección 8.3.2 para la estimación de parámetros.

pero ésta es la expansión binomial de $\left[\frac{1}{p} - \frac{(1-p)e^t}{p}\right]^{-k}$; por lo tanto, la función generadora de momentos está dada por:

$$m_X(t) = \frac{p^k}{[1 - (1 - p)e^t]^k}. \tag{4.43}$$

Con las distribuciones binomial, de Poisson, binomial negativa e hipergeométrica, se ha hecho un intento para proporcionar al lector distribuciones discretas de probabilidad que han demostrado ser modelos adecuados para muchos fenómenos interesantes y útiles de manera práctica. A pesar de que estas distribuciones son similares entre sí, cada una de ellas posee características distintas que brindan al usuario la información necesaria para una selección apropiada. También debe notarse que si un fenómeno no presenta todas las propiedades de una distribución determinada es suficiente para excluirla como modelo de probabilidad adecuado para ese fenómeno aleatorio.

Las distribuciones binomial, de Poisson y binomial negativa involucran ensayos de Bernoulli en el muestreo que se lleva a cabo con reemplazo. En la distribución binomial el muestreo se lleva a cabo con un número fijo de ensayos que tienen una probabilidad de éxito o fracaso constante. En la distribución de Poisson el número de ensayos es de tal manera infinito que la ocurrencia o no de un evento es constante en el tiempo y en el espacio. En la distribución binomial negativa, el muestreo se continúa hasta observar un determinado número de éxitos y el número de ensayos puede ser infinito. Por lo tanto, esta distribución es una alternativa factible de la de Poisson cuando la frecuencia de ocurrencia no es constante en el tiempo y el espacio. En la distribución hipergeométrica los ensayos no son independientes puesto que el muestreo se lleva a cabo sin reemplazo. No sólo el tamaño de la muestra es fijo, sino que se supone que la población es finita y, muchas veces, relativamente pequeña.

Referencias

1. A. G. Arbous and J. E. Kerrich, *Accident statistics and the concept of accident proneness*, Biometrics 7 (1951), 340–432.
2. N. L. Johnson and S. Kotz, *Discrete distributions*, Houghton Mifflin, Boston, 1969.
3. N. L. Johnson and F. C. Leone, *Statistics and experimental design*, Vol. I, Wiley, New York, 1977.
4. G. L. Lieberman and D. B. Owen, *Tables of the hypergeometric probability distribution*, Stanford Univ. Press, Stanford, Calif., 1961.
5. R. Pollard, *Collegiate football scores and the negative binomial distribution*, J. Amer. Statistical Assoc., **68** (1973), 351–352.
6. E. Williamson and M. H. Bretherton, *Tables of the negative binomial probability distribution*, Wiley, New York, 1963.

Ejercicios

4.1. Sea X una variable aleatoria con distribución binomial y parámetros n y p. Mediante la función de probabilidad binomial, verificar que $p(n - x; n, 1 - p) = p(x; n, p)$.

4.2. En una distribución binomial, sea X el número de éxitos obtenidos en diez ensayos donde la probabilidad de éxito en cada uno es de 0.8. Con el resultado del problema anterior, demostrar que la probabilidad de lograr de manera exacta seis éxitos es igual a la probabilidad de tener cuatro fracasos.

4.3. Mediante el empleo de la función de probabilidad binomial, verificar la siguiente fórmula de recursión:

$$p(x + 1; n, p) = \frac{(n - x)p}{(x + 1)(1 - p)} p(x; n, p).$$

4.4. Sea X una variable aleatoria con distribución binomial y parámetros $n = 8$ y $p = 0.4$. Emplear la fórmula de recursión del problema anterior para obtener las probabilidades puntuales de los valores de X. Hacer una gráfica de la función de probabilidad.

4.5. Sea X una variable aleatoria distribuida binomialmente con $n = 10$ y $p = 0.5$.

a) Determinar las probabilidades de que X se encuentre dentro de una desviación estándar de la media y a dos desviaciones estándares de la media.
b) ¿Cómo cambiarían las respuestas de *a)* si $n = 15$ y $p = 0.4$?

4.6. Supóngase que la probabilidad de tener una unidad defectuosa en una línea de ensamble es de 0.05. Si el número de unidades terminadas constituye un conjunto de ensayos independientes:

a) ¿Cuál es la probabilidad de que entre 20 unidades dos se encuentren defectuosas?
b) ¿Cuál es la probabilidad de que entre 20 unidades, dos como límite se encuentren defectuosas?
c) ¿Cuál es la probabilidad de que por lo menos una se encuentre defectuosa?

4.7. En una fábrica de circuitos electrónicos, se afirma que la proporción de unidades defectuosas de cierto componente que ésta produce, es del 5%. Un buen comprador de estos componentes revisa 15 unidades seleccionadas al azar y encuentra cuatro defectuosas. Si la compañía se encuentra en lo correcto y prevalecen las suposiciones para que la distribución binomial sea el modelo de probabilidad adecuado para esta situación, ¿cuál es la probabilidad de este hecho? Con base en el resultado anterior ¿puede concluirse que la compañía está equivocada?

4.8. La probabilidad de que un satélite, después de colocarlo en órbita, funcione de manera adecuada es de 0.9. Supóngase que cinco de éstos se colocan en órbita y operan de manera independiente:

a) ¿Cuál es la probabilidad de que, por lo menos, el 80% funcione adecuadamente?
b) Responder a *a)* si $n = 10$
c) Responder a *a)* si $n = 20$
d) ¿Son inesperados estos resultados? ¿Por qué?

4.9. Con base en encuestas al consumidor se sabe que la preferencia de éste con respecto a dos marcas, A y B, de un producto dado, se encuentra muy pareja. Si la opción de

compra entre estas marcas es independiente, ¿cuál es la probabilidad de que entre 25 personas seleccionadas al azar, no más de diez tengan preferencia por la marca A?

4.10. Supóngase que un examen contiene 15 preguntas del tipo falso o verdadero. El examen se aprueba contestando correctamente por lo menos nueve preguntas. Si se lanza una moneda para decidir el valor de verdad de cada pregunta, ¿cuál es la probabilidad de aprobar el examen?

4.11. Un vendedor de seguros sabe que la oportunidad de vender una póliza es mayor mientras más contactos realice con clientes potenciales. Si la probabilidad de que una persona compre una póliza de seguro después de la visita, es constante e igual a 0.25, y si el conjunto de visitas constituye un conjunto independiente de ensayos, ¿cuántos compradores potenciales debe visitar el vendedor para que la probabilidad de vender por lo menos una póliza sea de 0.80?

4.12. El gerente de un restaurante que sólo da servicio mediante reservación sabe, por experiencia, que el 15% de las personas que reservan una mesa no asistirán. Si el restaurante acepta 25 reservaciones pero sólo dispone de 20 mesas, ¿cual es la probabilidad de que a todas las personas que asistan al restaurante se les asigne una mesa?

4.13. Mediante la probabilidad de Poisson, demostrar la siguiente fórmula de recursión:

$$p(x + 1; \lambda) = \frac{\lambda}{(x + 1)} p(x; \lambda).$$

4.14. Sea X una variable aleatoria de Poisson con parámetro $\lambda = 2$. Emplear la fórmula del problema anterior para determinar las probabilidades puntuales de $X = 0, 1, 2, 3, 4, 5, 6, 7$ y 8, y hágase una gráfica de la función de probabilidad.

4.15. Para un volumen fijo, el número de células sanguíneas rojas es una variable aleatoria que se presenta con una frecuencia constante. Si el número promedio para un volumen dado es de nueve células para personas normales, determinar la probabilidad de que el número de células rojas para una persona se encuentra dentro de una desviación estándar del valor promedio y a dos desviaciones estándar del promedio.

4.16. El número de clientes que llega a un banco es una variable aleatoria de Poisson. Si el número promedio es de 120 por hora, ¿cuál es la probabilidad de que en un minuto lleguen por lo menos tres clientes? ¿Puede esperarse que la frecuencia de llegada de los clientes al banco sea constante en un día cualquiera?

4.17. Supóngase que en un cruce transitado ocurren de manera aleatoria e independiente dos accidentes por semana. Determinar la probabilidad de que ocurra un accidente en una semana y de que ocurran tres, en la semana siguiente.

4.18. Sea X una variable aleatoria binomial. Para $n = 20$, calcular las probabilidades puntuales binomiales y compararlas con las correspondientes probabilidades de Poisson para $p = 0.5, 0.3, 0.1$ y 0.01.

4.19. Una compañía compra cantidades muy grandes de componentes electrónicos. La decisión para aceptar o rechazar un lote de componentes se toma con base en una muestra aleatoria de 100 unidades. Si el lote se rechaza al encontrar tres o más unidades defectuosas en la muestra, ¿cuál es la probabilidad de rechazar un lote si éste contiene un 1% de componentes defectuosos? ¿Cuál es la probabilidad de rechazar un lote que contenga un 8% de unidades defectuosas?

4.20. El número de componentes que fallan antes de cumplir 100 horas de operación es una variable aleatoria de Poisson. Si el número promedio de éstas es ocho:

a) ¿Cuál es la probabilidad de que falle un componente en 25 horas?
b) ¿Cuál es la probabilidad de que fallen no más de dos componentes en 50 horas?
c) ¿Cuál es la probabilidad de que fallen por lo menos diez en 125 horas?

4.21. Mediante estudios recientes se ha determinado que la probabilidad de morir por causa de cierta vacuna contra la gripe es de 0.00002. Si se administra la vacuna a 100 mil personas y se supone que éstas constituyen un conjunto independiente de ensayos, ¿cuál es la probabilidad de que mueran no más de dos personas a causa de la vacuna?

4.22. Un fabricante asegura a una compañía que el porcentaje de unidades defectuosas es de sólo dos. La compañía revisa 50 unidades seleccionadas aleatoriamente y encuentra cinco defectuosas. ¿Qué tan probable es este resultado si el porcentaje de unidades defectuosas es el que el fabricante asegura?

4.23. El número de accidentes graves en una pla..ta i·:dustrial es de diez por año, de manera tal que el gerente instituye un plan que considera efectivo para reducir el número de accidentes en la planta. Un año después de ponerlo en marcha, sólo han ocurrido cuatro accidentes. ¿Qué probabilidad hay de cuatro o menos accidentes por año, si la frecuencia promedio aún es diez? Después de lo anterior, ¿puede concluirse que, luego de un año, el número de accidentes promedio ha disminuido?

4.24. El Departamento de Protección del Ambiente ha adquirido 40 instrumentos de precisión para medir la contaminación del aire en distintas localidades. Se seleccionan aleatoriamente ocho instrumentos y se someten a una prueba para encontrar defectos. Si cuatro de los 40 instrumentos se encuentran defectuosos, ¿cuál es la probabilidad de que la muestra contenga no más de un instrumento defectuoso?

4.25. Se sospecha que por causa de un error humano se han incluido en un embarque de 50 unidades, dos (o más) defectuosas. El fabricante admite el error y envía al cliente sólo 48 unidades. Antes de recibir el embarque, el cliente selecciona aleatoriamente cinco unidades y encuentra una defectuosa. ¿Debe reclamar una indemnización al fabricante?

4.26. Los jurados para una corte federal de distrito se seleccionan de manera aleatoria entre la lista de votantes del distrito. En un determinado mes se selecciona una lista de 25 candidatos. Ésta contiene los nombres de 20 hombres y cinco mujeres.

a) Si la lista de votantes se encuentra igualmente dividida por sexo. ¿cuál es la probabilidad de tener una lista que contenga a 20 hombres y cinco mujeres?
b) Supóngase que de esta lista se elige un jurado de doce personas, de las cuales sólo una es mujer. ¿Cuál es la probabilidad de este hecho, si los miembros del jurado se seleccionan de manera aleatoria?
c) Si el lector fuera el abogado de la defensa, ¿que podría argumentar mediante el empleo de las respuestas de las partes *a* y *b*?

4.27. Una compañía recibe un lote de 1 000 unidades. Para aceptarlo se seleccionan diez unidades de manera aleatoria, y se inspeccionan. Si ninguna se encuentra defectuosa, el lote se acepta; de otro modo, se rechaza. Si el lote contiene un 5% de unidades defectuosas:

a) Determinar la probabilidad de aceptarlo mediante el empleo de la distribución hipergeométrica.

b) Aproximar la respuesta de la parte *a* mediante el empleo de la distribución binomial.
c) Aproximar la respuesta de la parte *b* mediante el empleo de la distribución de Poisson.

4.28. En el ejercicio anterior, ¿cómo cambiarían las respuestas de las partes *a, b* y *c* si el tamaño del lote fuera de 40 unidades?

4.29. Considérese las funciones de probabilidad binomial y binomial negativa dadas por las expresiones 4.1 y 4.34, respectivamente. Demostrar que:

$$p_{NB}(x; k, p) = \frac{k}{x + k}\, p_B(k; x + k, p).$$

4.30. Sea X una variable aleatoria binomial negativa con parámetros $k = 3$ y $p = 0.4$. Emplee el resultado del problema anterior para calcular las probabilidades puntuales para los siguientes valores de X: 0, 1, 2, 3, 4 y 5.

4.31. Greenwood y Yule* dieron a conocer el número de accidentes ocurridos entre 414 operadores de maquinaria, en un periodo de tres meses consecutivos. En la tabla 4.10 la primera columna indica el número de accidentes sufridos por un mismo operador, y la segunda indica la frecuencia relativa para aquellos que habían sufrido la cantidad de accidentes indicada en el lapso de tres meses.

TABLA 4.10

x	*Frecuencia relativa*
0	0.715
1	0.179
2	0.063
3	0.019
4	0.010
5	0.010
6	0.002
7	0.000
8	0.002

Con el procedimiento del ejemplo 4.10, comparar las frecuencias relativas observadas con las correspondientes probabilidades si el número de accidentes es una variable aleatoria binomial negativa.

4.32. Un contador recientemente graduado pretende realizar el examen CPA. Si el número de veces que se hace el examen constituye un conjunto de eventos independientes con una probabilidad de aprobar igual a 0.6, ¿cuál es la probabilidad de que no se necesiten más de cuatro intentos para aprobar el examen? ¿Son válidas las suposiciones de independencia y probabilidad constante?

* *Encuesta acerca de la distribución representativa de la frecuencia de múltiples eventos, con especial referencia a la ocurrencia de múltiples ataques de enfermedades o accidentes repetidos,* J. of the Royal Statistical Soc. **83** (1920), 255.

4.33. En un departamento de control de calidad se inspeccionan las unidades terminadas que provienen de una línea de ensamble. Se piensa que la proporción de unidades defectuosas es de 0.05.

a) ¿Cuál es la probabilidad de que la vigésima unidad inspeccionada sea la segunda que se encuenre defectuosa?
b) Supóngase que la décimo quinta unidad inspeccionada es la segunda que se encuentra defectuosa. ¿Cuál es la probabilidad de este hecho bajo condiciones determinadas?

4.34. De las distribuciones binomial, Poisson, hipergeométrica y binomial negativa, ¿cuáles no consideraría si alguien le dijera, de una distribución en particular que:

a) ¿La media es igual a la varianza?
b) ¿La media es más grande que la varianza?
c) ¿La media es menor que la varianza?
d) El tercer momento, alrededor de la media, ¿es negativo?
e) ¿El fenómeno aleatorio de interés constituye un grupo de ensayos independientes?
f) ¿El muestreo se lleva a cabo con reemplazo?
g) ¿El muestreo se lleva a cabo sin reemplazo?

APÉNDICE

Deducción de la función de probabilidad de Poisson

Sea $p(x; t)$ la probabilidad de tener, de manera exacta, X ocurrencias en un intervalo t, y supóngase lo siguiente:

1. En este intervalo, los eventos ocurren de manera independiente.
2. La probabilidad de una sola ocurrencia, en un intervalo muy pequeño dt es νdt, en donde ν es la frecuencia constante de ocurrencia y $(\nu > 0)$.
3. El intervalo dt es tan pequeño, que la probabilidad de tener más de una ocurrencia en dt es despreciable.

El evento que en el tiempo $t + dt$ ha ocurrido exactamente x veces, puede llevarse a cabo de dos maneras diferentes y excluyentes:

1. Existen x ocurrencia por tiempo t, con probabilidad $p(x; t)$ y ninguna en dt, con probabilidad $(1 - \nu dt)$. Dada la suposición de independencia, la probabilidad conjunta es $p(x; t)(1 - \nu dt)$.
2. Existen $x - 1$ ocurrencias por tiempo t, con probabilidad $p(x - 1; t)$ y una durante dt, con probabilidad νdt. Otra vez, dada la suposición de independencia, la probabilidad conjunta es: $p(x - 1; t)\nu dt$.

Esto es:

$$p(x; t + dt) = p(x; t)(1 - \nu dt) + p(x - 1; t)\nu dt.$$

Después de multiplicar, transportar $p(x;\, t)$ al primer miembro, y dividir por dt, se tiene:

$$\frac{p(x;\, t + dt) - p(x;\, t)}{dt} = \nu[p(x - 1;\, t) - p(x;\, t)].$$

Si se toma el límite conforme $dt \to 0$, por definición se tiene:

$$\frac{dp(x;\, t)}{dt} = \nu[p(x - 1;\, t) - p(x;\, t)], \tag{4.44}$$

que es una ecuación diferencial lineal con respecto a t y una ecuación de diferencias finitas de primer orden, con respecto a x. Si $x = 0$, la ecuación (4.44) se convierte en

$$\frac{dp(0;\, t)}{dt} = \nu[p(-1;\, t) - p(0;\, t)]$$

$$= -\nu p(0;\, t),$$

dado que $p(-1;t)$ tiene que ser cero. La solución general de la ecuación diferencial lineal

$$\frac{dp(0;\, t)}{dt} = -\nu p(0;\, t)$$

se obtiene mediante separación de variables e integración en ambos miembros, lo que da como resultado:

$$\ln[p(0;\, t)] = \ln(c) - \nu t,$$

o

$$p(0;\, t) = ce^{-\nu t}$$

Dado que la probabilidad de tener cero ocurrencias en un intervalo $t = 0$, debe ser 1, $c = 1$, y

$$p(0;\, t) = e^{-\nu t}.$$

Si $x = 1$, (4.44) se convierte en

$$\frac{dp(1;\, t)}{dt} = \nu[p(0;\, t) - p(1;\, t)],$$

o

$$\frac{dp(1;\, t)}{dt} + \nu p(1;\, t) = \nu e^{-\nu t} \tag{4.45}$$

La ecuación (4.45) es una ecuación diferencial no homogénea con la condición inicial de que $p(1;\, 0) = 0$ dado que la probabilidad de tener exactamente una

ocurrencia en $t = 0$ debe ser cero. La solución de (4.45) es

$$p(1; t) = (\nu t)e^{-\nu t}$$

De manera similar, para $x = 2$ y $p(2; 0) = 0$, (4.44) se reduce a

$$\frac{dp(2; t)}{dt} + \nu p(2; t) = \nu^2 t e^{-\nu t},$$

cuya solución es

$$p(2; t) = \frac{(\nu t)^2 e^{-\nu t}}{2!}.$$

Al continuar este proceso puede deducirse que la probabilidad de tener exactamente x ocurrencias en t es

$$p(x; t) = \frac{(\nu t)^x e^{-\nu t}}{x!}, \qquad x = 0, 1, 2, \ldots \tag{4.46}$$

siempre que $p(x; 0) = 0$. Si se sustituye $\lambda = \nu t$ en (4.46), el resultado es la función de probabilidad de Poisson.

APÉNDICE

Demostración del teorema 4.1

Al multiplicar numerador y denominador por n^x y sustituir $n!/(n - x)! = n(n - 1)(n - 2) \cdots (n - x + 1)$, la función de probabilidad binomial es:

$$\begin{aligned} p(x; n, p) &= \frac{n(n-1)(n-2)\cdots(n-x+1)}{n^x x!}(np)^x(1-p)^{n-x} \\ &= \frac{n(n-1)(n-2)\cdots(n-x+1)}{n^x}\frac{\lambda^x}{x!}(1-p)^{n-x} \\ &= 1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\frac{\lambda^x}{x!}(1-p)^{n-x} \\ &= \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)}{(1-p)^x}\frac{\lambda^x}{x!}(1-p)^n. \end{aligned} \tag{4.47}$$

Dado que:

$$(1-p)^n = [(1-p)^{-1/p}]^{-np} = [(1-p)^{-1/p}]^{-\lambda},$$

y por definición:

$$\lim_{z \to 0} (1 + z)^{1/z} = e,$$

mediante el cambio de variable $z = -p$, se tiene

$$\lim_{p \to 0} [(1 - p)^{-1/p}]^{-\lambda} = e^{-\lambda}.$$

Además,

$$\lim_{n \to \infty} \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{x - 1}{n}\right) = 1$$

y

$$\lim_{p \to 0} (1 - p)^x = 1.$$

Al sustituir en (4.47),

$$\lim_{\substack{n \to \infty \\ p \to 0}} p(x; n, p) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots .$$

CAPÍTULO CINCO

Algunas distribuciones continuas de probabilidad

5.1 Introducción

Estas distribuciones se emplearon en el estudio de fenómenos aleatorios en disciplinas como la ingeniería y las ciencias aplicadas o bien los negocios y la economía. En este capítulo se desarrollará un método para determinar la distribución de probabilidad de una función de variable aleatoria y se introducirán los conceptos básicos para la generación, por computadora, de números aleatorios.

De manera específica se estudiarán los siguientes modelos de probabilidad: normal, uniforme, beta, gama, de Weibull y exponencial negativa. La forma de abordar los temas será la misma que se empleó en el capítulo cuatro. Se discutirán las propiedades de cada modelo y se indicarán áreas de aplicación específica, con lo que se pretende proporcionar al lector una idea y comprensión suficiente para utilizar los modelos de manera apropiada.

5.2 La distribución normal

La *distribución normal* o *Gausiana* es indudablemente la más importante y la de mayor uso de todas las distribuciones continuas de probabilidad. Es la piedra angular en la aplicación de la inferencia estadística en el análisis de datos, puesto que las distribuciones de muchas estadísticas muestrales tienden hacia la distribución normal conforme crece el tamaño de la muestra. La apariencia gráfica de la distribución normal es una curva simétrica con forma de campana, que se extiende sin límite tanto en la dirección positiva como en la negativa. Un gran número de estudios indica que la distribución normal proporciona una adecuada representación, por lo menos en una primera aproximación, de las distribuciones de una gran cantidad de variables físicas. Algunos ejemplos específicos incluyen datos meteorológicos tales como la temperatura y la precipitación pruvial, mediciones efectuadas en organismos vivos, calificaciones en pruebas de actitud, mediciones físicas de partes manu-

facturadas, errores de instrumentación y otras desviaciones de las normas establecidas, etc. Sin embargo, debe tenerse mucho cuidado al suponer para una situación dada un modelo de probabilidad normal sin previa comprobación. Si bien es cierto que la distribución normal es la que tiene un mayor uso, es también de la que más se abusa. Quizá esto se deba a la mala interpretación de la palabra "normal", especialmente si se aplica su significado literal de "patrón o estándar aceptado". Suponer de manera errónea una distribución normal puede llevar a errores muy serios. Es posible que una distribución normal proporcione de manera razonable una buena aproximación alrededor de la media de una variable aleatoria; sin embargo, puede resultar para valores extremos que se encuentren en cualquier dirección. Por ejemplo, si se diseña cierto material para resistir una cantidad dada de presión, que se supone se encuentra distribuida normalmente alrededor de un valor promedio, y el diseño se hace con base en esta suposición, el material puede verse seriamente dañado al aplicársele una presión muy elevada.

En la definición 5.1 se proporciona la función de densidad de probabilidad de la distribución normal, la cual fue descubierta por DeMoivre en 1733 como una forma límite de la función de probabilidad binomial; después la estudió Laplace. También se conoce como distribución Gausiana porque Gauss la citó en un artículo que publicó en 1809. Durante el siglo XIX se empleó de manera extensa por científicos que habían notado que los errores, al llevar a cabo mediciones físicas, frecuentemente seguían un patrón que sugería la distribución normal.

Definición 5.1 Se dice que una variable aleatoria X se encuentra normalmente distribuida si su función de densidad de probabilidad está dada por

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \qquad \begin{array}{l} -\infty < x < \infty \\ -\infty < \mu < \infty, \sigma > 0. \end{array} \tag{5.1}$$

Los parámetros de la distribución normal son μ y σ y además determinan de manera completa la función de densidad de probabilidad. Como se verá posteriormente, estos parámetros son la media y la desviación estándar de X, respectivamente. En la figura 5.1 se proporcionan varias gráficas de (5.1) para distintos valores de μ a σ fijo y viceversa.

Es obvio que para cualquier par de valores μ y σ, (5.1) es simétrica y tiene forma de campana. Si se obtienen las dos primeras derivadas de $f(x;\ \mu,\ \sigma)$ con respecto a x y se igualan a cero, se tiene que el valor máximo de $f(x;\ \mu, \sigma)$ ocurre cuando $x = \mu$, y los valores $x = \mu \pm \sigma$ son las abcisas de los dos puntos de inflexión de la curva. En un apéndice al final de este capítulo se proporciona la demostración de que (5.1) es una función de densidad de probabilidad.

La media de una variable aleatoria distribuida normalmente se encuentra definida por:

$$E(X) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} x \exp[-(x-\mu)^2/2\sigma^2]dx. \tag{5.2}$$

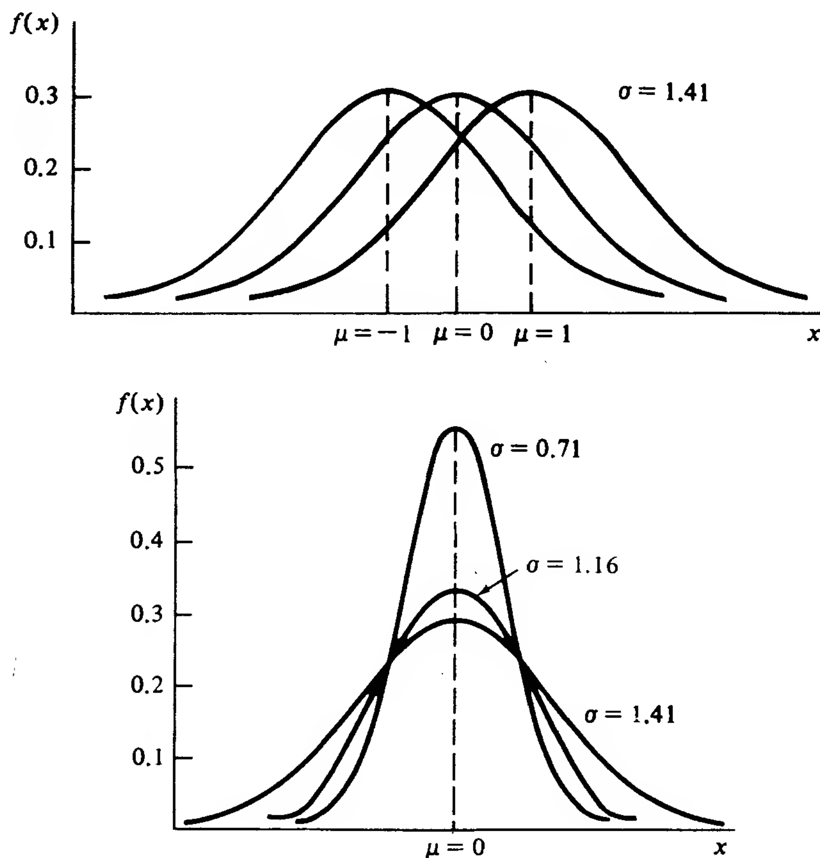


FIGURA 5.1 Gráficas de la función de densidad normal para diferentes valores de μ y σ

Se pretende demostrar que $E(X) = \mu$. Supóngase que a (5.2) se suma y se resta

$$\frac{\mu}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty}\exp[-(x-\mu)^2/2\sigma^2]dx.$$

La identidad se mantiene, pero después de reacomodar términos se tiene

$$\begin{aligned} E(X) &= \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty}(x-\mu)\exp[-(x-\mu)^2/2\sigma^2]dx \\ &\quad + \frac{\mu}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty}\exp[-(x-\mu)^2/2\sigma^2]dx \\ &= \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty}(x-\mu)\exp[-(x-\mu)^2/2\sigma^2]dx + \mu, \end{aligned} \qquad (5.3)$$

dado que el valor de la segunda integral es uno. Al efectuar un cambio de variable de integración en (5.3) de manera tal que $y = (x - \mu)/\sigma$, $x = \sigma y + \mu$, y $dx = \sigma\, dy$, se tiene:

$$E(X) = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y \exp(-y^2/2)dy + \mu$$

$$= -\frac{\sigma}{\sqrt{2\pi}} \exp(-y^2/2) \Big|_{-\infty}^{\infty} + \mu = \mu. \tag{5.4}$$

El lector recordará de sus cursos de cálculo que la última integral es cero porque el integrando es una función impar* y la integración se lleva a cabo sobre un intervalo simérico alrededor de cero.

Una distribución normal es simétrica alrededor de su media μ. Si el valor máximo de la función de densidad de probabilidad normal ocurre cuando $x = \mu$, μ es la media, la mediana y la moda de cualquier variable aleatoria distribuida normalmente.

Para encontrar los demás momentos, se determinará la función generadora de momentos. Por definición:

$$m_{X-\mu}(t) = E[e^{t(X-\mu)}] = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} \exp[t(x - \mu)]\exp[-(x - \mu)^2/2\sigma^2]dx$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2\sigma^2}[(x - \mu)^2 - 2\sigma^2 t(x - \mu)] \right\}dx.$$

Se completa el cuadrado en el interior del paréntesis rectangular y se tiene:

$$(x - \mu)^2 - 2\sigma^2 t(x - \mu) = (x - \mu)^2 - 2\sigma^2 t(x - \mu) + \sigma^4 t^2 - \sigma^4 t^2$$

$$= (x - \mu - \sigma^2 t)^2 - \sigma^4 t^2$$

y:

$$m_{X-\mu}(t) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} \exp(\sigma^2 t^2/2)\exp\{-[x - (\mu + \sigma^2 t)]^2/2\sigma^2\}dx$$

$$= \exp(\sigma^2 t^2/2) \cdot \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} \exp\{-[x - (\mu + \sigma^2 t)]^2/2\sigma^2\}dx$$

$$= \exp(\sigma^2 t^2/2), \tag{5.5}$$

dado que el integrando junto con el factor $1/\sqrt{2\pi}\,\sigma$ es una función de densidad de probabilidad normal con parámetros $\mu + \sigma^2 t$ y σ.

Al desarrollar (5.5) en serie de potencias se tiene:

$$m_{X-\mu}(t) = 1 + \frac{(\sigma t)^2}{2} + \frac{(\sigma t)^4}{4 \cdot 2!} + \frac{(\sigma t)^6}{8 \cdot 3!} + \frac{(\sigma t)^8}{16 \cdot 4!} + \cdots .$$

* Se dice que una función $f(x)$ es impar si $f(-x) = -f(x)$. Entonces $\int_{-a}^{a} f(x)dx = 0$. Se dice que una función $f(x)$ es par si $f(-x) = f(x)$. Entonces $\int_{-a}^{a} f(x)dx = 2\int_0^a f(x)dx$.

Cuando las potencias impares de t no se encuentran presentes, todos los momentos centrales de X de orden impar son cero, de esta forma se asegura la simetría de la curva.

La segunda derivada de $m_{X-\mu}(t)$ evaluada en $t = 0$ es la varianza y está dada por:

$$Var(X) = \left.\frac{d^2 m_{X-\mu}(t)}{dt^2}\right|_{t=0} = \sigma^2 + \frac{12t^2\sigma^4}{4\cdot 2!} + \frac{30t^4\sigma^6}{8\cdot 3!} + \cdots \Big|_{t=0} = \sigma^2; \quad (5.6)$$

de esta manera la desviación estándar es σ. De manera similar, la cuarta derivada de $m_{X-\mu}(t)$ evaluada en $t = 0$ es el cuarto momento central, el cual es:

$$\mu_4 = \left.\frac{d^4 m_{X-\mu}(t)}{dt^4}\right|_{t=0} = 3\sigma^4 + \frac{360t^2\sigma^6}{8\cdot 3!} + \cdots \Big|_{t=0} = 3\sigma^4 \quad (5.7)$$

De acuerdo con lo anterior, para cualquier distribución normal el coeficiente de asimetría es $\alpha_3(X) = 0$, mientras que la curtosis relativa es $\alpha_4(X) = 3\sigma^4/\sigma^4 = 3$. Para momentos alrededor del cero, puede determinarse la función generadora de momentos de X mediante el empleo directo de la función generadora de momentos centrales (o viceversa). Dado que

$$\begin{aligned} m_{X-\mu}(t) &= E[e^{t(X-\mu)}] \\ &= \exp(-\mu t)E[\exp(tX)] \\ &= \exp(-\mu t)m_X(t), \end{aligned}$$

para una distribución normal

$$\exp(-\mu t)m_X(t) = \exp(\sigma^2 t^2/2)$$

y

$$m_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right). \quad (5.8)$$

La probabilidad de que una variable aleatoria normalmente distribuida X sea menor o igual a un valor específico, x está dada por la función de distribución acumulativa

$$P(X \leq x) = F(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x} \exp[-(t-\mu)^2/2\sigma^2]dt. \quad (5.9)$$

La integral en (5.9) no puede evaluarse en forma cerrada; sin embargo, se puede tabular $F(x; \mu, \sigma)$ como una función de μ y σ, lo que necesitaría una tabla para cada par de valores. Como existe un número infinito de valores de μ y σ, esta tarea es virtualmente imposible. Afortunadamente, lo anterior puede simplificarse mediante el empleo de la siguiente transformación: sea Z una variable aleatoria definida por la siguiente relación:

$$Z = (X - \mu)/\sigma, \tag{5.10}$$

en donde μ y σ son la media y la desviación estándar de X, respectivamente. De acuerdo con lo anterior, Z* es una variable aleatoria estandarizada con media cero y desviación estándar uno, de acuerdo con lo que se discutió en el capítulo tres.

Si la transformación (5.10) se sustituye en (5.9), entonces:

$$P(X \leq x) = P[Z \leq (x - \mu)/\sigma] = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{(x-\mu)/\sigma} \exp(-z^2/2)(\sigma dz)$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{(x-\mu)/\sigma} \exp(-z^2/2)dz. \tag{5.11}$$

El integrando en (5.11) junto con el factor $1/\sqrt{2\pi}$ es la función de densidad de probabilidad de la variable aleatoria normal estandarizada Z. Esto es, si X se encuentra normalmente distribuida con media μ y desviación estándar σ, entonces $Z = (X - \mu)/\sigma$ también se encuentra normalmente distribuida con media cero y desviación estándar uno. Así, para $z = (x - \mu)/\sigma$, $P(X \leq x) = P(Z \leq z)$ y

$$F_X(x; \mu, \sigma) = F_Z(z; 0, 1), \tag{5.12}$$

donde $F_Z(z;\ 0,\ 1)$ es la función de distribución acumulativa de la función de probabilidad normal estandarizada. En la figura 5.2 se proporciona la gráfica de la función de distribución para la variable aleatoria normal estandarizada.

* Se empleará Z para denotar una variable aleatoria normal estandarizada.

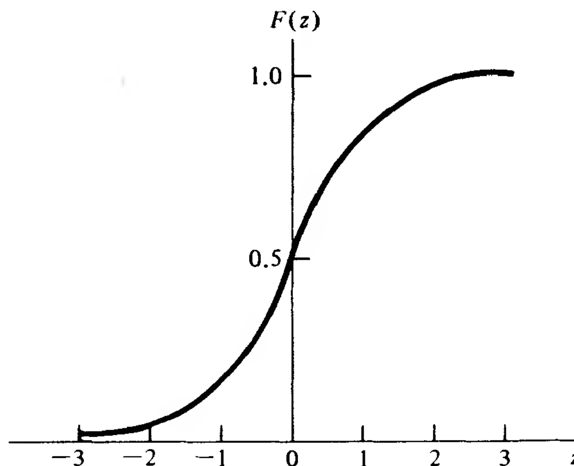


FIGURA 5.2 Función de distribución acumulativa de la normal estándar

La función $F_Z(z;\ 0,\ 1)$ se encuentra tabulada, de manera extensa, y se da en la tabla D del apéndice. Para cualquier valor específico de z, el correspondiente valor en la tabla es la probabilidad de que la variable aleatoria normal estándar Z sea menor o igual a z; esto es,

$$P(Z \le z) = F_Z(z; 0, 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-t^2/2)dt. \tag{5.13}$$

En este momento es conveniente introducir la notación $X \sim N(\mu,\ \sigma)$ para denotar que la variable X se encuentra distribuida normalmente con media μ y desviación estándar σ. En lo que sigue se examinará cómo puede determinarse la probabilidad de que un valor de X se encuentre entre a y b, si $X \sim N(\mu,\ \sigma)$. Por definición:

$$P(a \le X \le b) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{a}^{b} \exp[-(x - \mu)^2/2\sigma^2]dx,$$

pero, mediante el empleo de (5.3) se tiene:

$$\begin{aligned} P(a \le X \le b) &= P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} \exp(-z^2/2)dz \\ &= F_Z\left(\frac{b-\mu}{\sigma}; 0, 1\right) - F_Z\left(\frac{a-\mu}{\sigma}; 0, 1\right). \end{aligned} \tag{5.14}$$

En otras palabras, la probabilidad de que X esté entre a y b es, de manera exacta, la misma probabilidad de que Z se encuentre entre $(a\ -\ \mu)/\sigma$ y $(b\ -\ \mu)/\sigma$, en donde Z es $N(0,\ 1)$. En la figura 5.3 se ilustra esta correspondencia de probabilidades.

Se ilustrará el empleo de la tabla D mediante los siguientes ejemplos.

Ejemplo 5.1 Si X es $N(\mu,\sigma)$, ¿cuáles son las probabilidades de que el valor de X se encuentre a una, dos y tres veces la desviación estándar de la media?

$$\begin{aligned} P(\mu - \sigma \le X \le \mu + \sigma) &= P\left(\frac{\mu - \sigma - \mu}{\sigma} \le Z \le \frac{\mu + \sigma - \mu}{\sigma}\right) \\ &= P(-1 \le Z \le 1) \\ &= F_Z(1; 0, 1) - F_Z(-1; 0, 1) \\ &= 0.6826. \\ P(\mu - 2\sigma \le X \le \mu + 2\sigma) &= P(-2 \le Z \le 2) \\ &= F_Z(2; 0, 1) - F_Z(-2; 0, 1) = 0.9544. \\ P(\mu - 3\sigma \le X \le \mu + 3\sigma) &= P(-3 \le Z \le 3) \\ &= F_Z(3; 0, 1) - F_Z(-3; 0, 1) = 0.9974. \end{aligned}$$

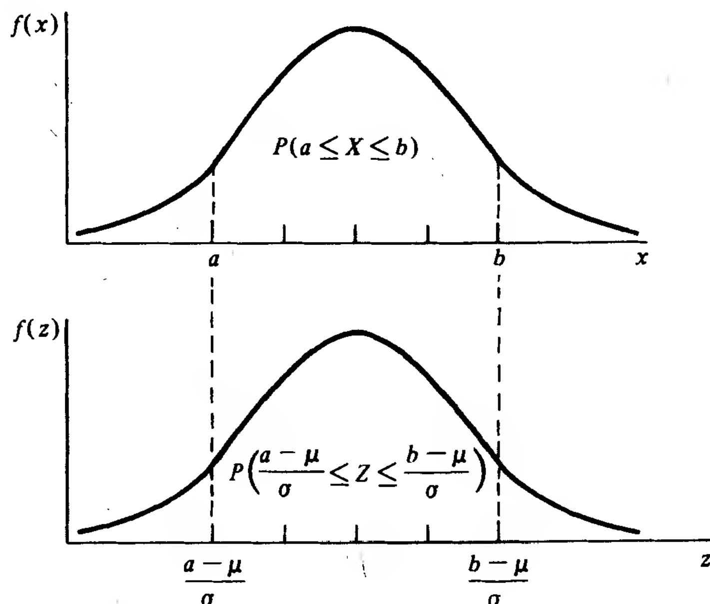


FIGURA 5.3 Correspondencia entre las probabilidades de X y de Z

Así, para cualquier variable aleatoria normal las probabilidades "una sigma", "dos sigma" y "tres sigma" son 0.6826, 0.9544 y 0.9974 respectivamente. Este resultado indica que para la distribución normal existe una gran concentración de valores alrededor de la media.

Ejemplo 5.2 Sea X una variable aleatoria que representa la inteligencia medida por medio de pruebas CI. Si X es $N(100, 10)$, obtener las probabilidades de que X sea mayor que 100, menor que 85, a lo más 112, por lo menos 108, más grande que 90, y entre 95 y 120.

Debe notarse que al resolver problemas de esta clase, el lector puede encontrar de gran ayuda graficar las correspondientes áreas bajo las curvas de densidad normal, como se ilustra en la figura 5.3. Dado que la distribución de probabilidad de X es simétrica alrededor de su media, la probabilidad de que X sea mayor que este valor es, por definición, 0.5. Las otras probabilidades se obtienen de la siguiente forma:

$$P(X < 85) = P\left(Z < \frac{85 - 100}{10}\right) = P(Z < -1.5)$$

$$= F_Z(-1.5; 0, 1) = 0.0668.$$

$$P(X \leq 112) = P(Z \leq 1.2) = F_Z(1.2; 0, 1) = 0.8849.$$

$$P(X \geq 108) = P(Z \geq 0.8) = 1 - F_Z(0.8; 0, 1) = 0.2119.$$

$$P(X > 90) = P(Z > -1) = 1 - F_Z(-1; 0, 1) = 0.8413.$$

$$P(95 \leq X \leq 120) = P(-0.5 \leq Z \leq 2) = F_Z(2; 0, 1) - F_Z(-0.5; 0, 1) = 0.6687.$$

Ejemplo 5.3 Supóngase que la demanda mensual de cierto producto se encuentra aproximada por una variable aleatoria normal con media de 200 y desviación estándar igual a 40 unidades. ¿Qué tan grande debe ser el inventario disponible a principio de un mes para que la probabilidad de que la existencia se agote no sea mayor de 0.05?

Sea X la demanda mensual, entonces X es $N(200, 40)$. Lo que se desea obtener es el valor del cuantil $x_{0.95}$ para el nivel de inventario a principio del mes, de manera tal que la probabilidad de que la demanda exceda a $x_{0.95}$ (existencias agotada) no sea mayor de 0.05. Esto es:

$$P(X > x_{0.95}) = 0.05$$

o

$$P(X \leq x_{0.95}) = 0.95.$$

De lo anterior se sigue que:

$$P[Z \leq (x_{0.95} - 200)/40] = 0.95$$

o

$$P(Z \leq z_{0.95}) = F_Z(z_{0.95}; 0, 1) = 0.95,$$

donde $z_{0.95} = (x_{0.95} - 200)/40$ es el valor cuantil correspondiente a la variable aleatoria normal estándar. Para obtener $z_{0.95}$ de la tabla D, primero se busca la probabilidad más cercana a 0.95. Una vez que se encuentra este valor, se toman los correspondientes valores del renglón y la columna y se interpola para encontrar el valor deseado de $z_{0.95}$. Por ejemplo, $z_{0.95}$ tiene un valor aproximado de 1.645 y dado que $z_{0.95} = (x_{0.95} - 200)/40$, $x_{0.95}$ tiene un valor de 265.8. Esto significa que el inventario a principio de cada mes no debe ser menor de 266 unidades para que la probabilidad de agotar las existencias no sea mayor de 0.05.

Ejemplo 5.4 Supóngase que el diámetro externo de cierto tipo de cojinetes se encuentra, de manera aproximada, distribuido normalmente con media igual a 3.5 cm y desviación estándar igual a 0.02 cm. Si el diámetro de estos cojinetes no debe ser menor de 3.47 cm ni mayor de 3.53 cm, ¿cuál es el porcentaje de cojinetes, durante el proceso de su manufactura, que debe desecharse?

Sea X el diámetro del cojinete, en donde X es $N(3.5, 0.02)$. La probabilidad de que el diámetro se encuentre entre 3.47 cm y 3.53 es:

$$\begin{aligned} P(3.47 \le X \le 3.53) &= P\left(\frac{3.47 - 3.5}{0.02} \le Z \le \frac{3.53 - 3.5}{0.02}\right) \\ &= P(-1.5 \le Z \le 1.5) \\ &= F_Z(1.5; 0, 1) - F_Z(-1.5; 0, 1) \\ &= 0.8664. \end{aligned}$$

Dado que el 86.64% de los cojinetes cumplen con las especificaciones determinadas, se deduce que $1 - 0.8664 = 0.1336$, o, en otras palabras, debe desecharse el 13.36% de la producción.

En el ejemplo 3.11 se determinó que para la distribución normal estándar los valores del primero y tercer cuantil son, de manera aproximada, iguales a -0.675 y 0.675 mientras que los correspondientes a los deciles primero y noveno son alrededor de -1.28 y 1.28 respectivamente. De (5.10) se sigue que si X es $N(\mu,\sigma)$, los valores de los cuantiles primero y tercero de X son $x_{0.25} = -0.675\sigma + \mu$ y $x_{0.75} = 0.675\sigma + \mu$. De esta manera el recorrido intercuantil es $x_{0.75} - x_{0.25} = 1.35\sigma$. De manera similar, los valores de los deciles primero y noveno son: $x_{0.10} = -1.28\sigma + \mu$ y $x_{0.90} = 1.28\sigma + \mu$, y el recorrido interdecil está dado por $x_{0.90} - x_{0.10} = 2.56\sigma$. Del ejemplo 3.11, se puede concluir que si $X \sim N(\mu,\sigma)$, la desviación media de X es

$$E|X - \mu| = 0.7979\sigma. \tag{5.15}$$

La tabla 5.1 contiene las propiedades básicas de la distribución normal.

Ejemplo 5.5 La primera columna de la tabla 5.2 contiene los intervalos de respuestas correctas para la prueba de matemáticas (SAT); la segunda, el correspondiente número de calificaciones observadas para el periodo 1979-1980, tal y como fueron dadas a conocer en el *College Board ATP Summary Report;* la tercera columna, las frecuencias relativas, las restantes, información con respecto a si las calificaciones para la prueba SAT obtenidas por los hombres estaban distribuidas normalmente con media 491* y desviación estándar igual a 120*.

* Estos datos se proporcionan en el College Board ATP Summary Report, 1979-1980.

TABLA 5.1 Propiedades básicas de la distribución normal

Función de densidad de probabilidad	*Parámetros*
$f(x; \mu, \sigma) = \dfrac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\dfrac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2\right]$, $-\infty < x < \infty$	μ, $-\infty < \mu < \infty$; σ, $\sigma > 0$

Media	*Varianza*	*Desviación media*	*Recorrido intercuantil*	*Recorrido interdecil*	*Coeficiente de asimetría*	*Curtosis relativa*
μ	σ^2	0.7979σ	1.35σ	2.56σ	0	3

TABLA 5.2 Calificaciones obtenidas en la prueba de matemáticas SAT por los estudiantes del tercer año de preparatoria en el ciclo 1979-1980

Número de respuestas correctas	*Número de exámenes*	*Frecuencia relativa*	*Intervalo normal estándar*	*Probabilidad del intervalo*	*Número esperado*
(200–249)	3 423	0.0072	(−2.425– −2.01)	0.0146	6 981.62
(250–299)	18 434	0.0385	(−2.01– −1.59)	0.0337	16 115.10
(300–349)	39 913	0.0835	(−1.59– −1.18)	0.0631	30 173.98
(350–399)	51 603	0.1079	(−1.18– −0.76)	0.1046	50 018.99
(400–449)	61 691	0.1290	(−0.76– −0.34)	0.1433	68 525.06
(450–499)	72 186	0.1510	(−0.34–0.075)	0.1630	77 945.46
(500–549)	72 804	0.1522	(0.075–0.49)	0.1580	75 554.49
(550–599)	58 304	0.1219	(0.49–0.91)	0.1307	62 499.83
(600–649)	46 910	0.0981	(0.91–1.325)	0.0888	42 463.54
(650–699)	30 265	0.0633	(1.325–1.74)	0.0517	24 722.58
(700–749)	16 246	0.0340	(1.74–2.16)	0.0255	12 193.92
(750–800)	6 414	0.0134	(2.16–2.575)	0.0104	4 973.21
Totales	478 193	1.0000		0.9874	472 167.78

Mientras que, de manera aparente, existe una similitud entre las frecuencias teóricas y las observadas, queda aún por contestar la pregunta acerca de cuándo puede rechazarse o no (véase Cap. 10) la hipótesis de que las calificaciones de la prueba SAT se distribuyeron normalmente con media 491 desviación estándar igual a 120. Como se mencionó, siempre es importante verificar lo que ocurre en los extremos de la distribución observada. Por ejemplo, se sabe que para la prueba SAT es imposible obtener calificaciones para los eventos $X < 200$ y $X > 800$. Sin embargo, si $X \sim N(491)$, las correspondientes probabilidades son 120), $P(X < 200) = 0.0075$ y $P(X > 800) = 0.005$. El siguiente ejemplo debe ilustrar de manera más clara la falta de concordancia en los extremos, entre las distribuciones observadas y teórica.

Ejemplo 5.6 El número de unidades de un cierto producto que un comerciante vende al día varía de manera aleatoria con cambios muy pequeños que se deben a la temporada o al día de la semana. Con base en información anterior, se cree que la demanda diaria de este producto es una variable aleatoria normal con media y desviación estándar iguales a 100 y 12 unidades, respectivamente. Para comprobar su grado de creencia, el vendedor anota la demanda diaria durante los últimos 102 días y la agrupa como se muestra en la tabla 5.3. Comparar las frecuencias relativas que se observaron con las frecuencias teóricas al suponer una distribución normal con media 100 y desviación estándar 12.

Como se ilustra en la figura 5.4, las frecuencias relativas que se observan en la demanda diaria sugieren una curva en forma de campana. Sin embargo, la tabla 5.4 en que se comparan las frecuencias relativas teórica y observada, muestra una discrepancia muy grande en los extremos a pesar de que existe una buena concordancia alrededor de la media. Suponer una distribución normal para este tipo de si-

TABLA 5.3 Demanda diaria de un producto

Demanda diaria	*Frecuencia*
(55–64)	6
(65–74)	4
(75–84)	6
(85–94)	20
(95–104)	32
(105–114)	18
(115–124)	6
(125–134)	6
(135–144)	4

tuación puede llevar a errores muy grandes cuando es necesario tener información sobre los extremos.

Recuérdese que la distribución binomial es una forma límite de la distribución de Poisson cuando n es grande y p pequeño. Se desea demostrar que la distribución normal es una forma límite de la binomial cuando n es grande y p no tiene un valor cercano a cero o a uno. El siguiente teorema, que se conoce como teorema del límite de DeMoivre-Laplace, asegura una aproximación adecuada mediante la distribución normal de las probabilidades binomiales si n es suficientemente grande.

Teorema 5.1 Sea X una variable aleatoria binomial con media np y desviación estándar $\sqrt{np(1-p)}$. La distribución de la variable aleatoria tiende a la normal

$$Y = \frac{X - np}{\sqrt{np(1 - p)}} \tag{5.16}$$

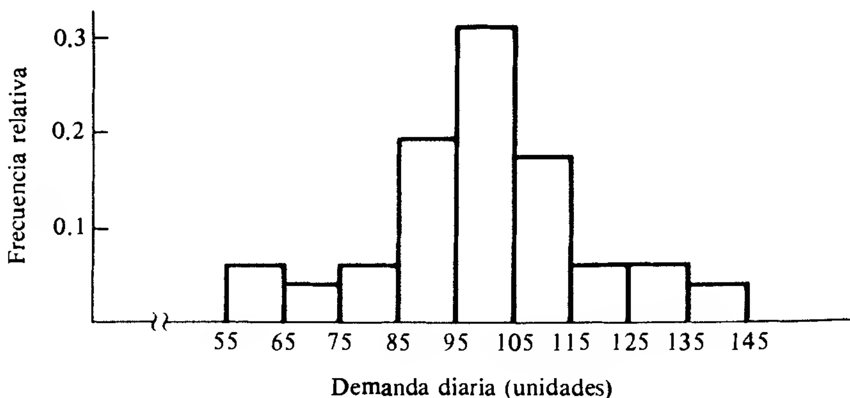


FIGURA 5.4 Frecuencias relativas que se observan para la demanda diaria de un producto

estándar conforme el número de ensayos independientes $n \to \infty$. Se proporciona un desarrollo de la prueba en un apéndice al final de este capítulo.

La esencia del teorema 5.1 es que si X es una variable aleatoria binomial, para la que el número de ensayos independientes es suficientemente grande, se dice que X posee una distribución normal aproximada con media np y desviación estándar $\sqrt{np(1-p)}$. De hecho, la aproximación es adecuada tanto como $np > 5$ cuando $p \leq 1/2$, o cuando $n(1-p) > 5$ para $p > 1/2$. Esto es,

$$P(a \leq X_B \leq b) \approx P\left(\frac{a - np}{\sqrt{np(1-p)}} \leq Z_N \leq \frac{b - np}{\sqrt{np(1-p)}}\right) \tag{5.17}$$

en donde Z_N es $N(0.1)$.

La aproximación dada por (5.17) puede mejorarse si se toma en cuenta que lo que se desea es aproximar probabilidades para una variable aleatoria discreta a partir del intervalo de probabilidades de una variable aleatoria continua. Por ejemplo, se desea determinar la probabilidad de que X tome un valor igual a x. Se sabe que para cualquier valor específico x de una variable aleatoria binomial, la probabilidad puntual es distinta de cero. Sin embargo, si se emplea la aproximación normal dada por el teorema 5.1, $P[Z = (x - np)/\sqrt{np(1-p)}] = 0$. En lugar de emplear la expresión anterior, se usará la aproximación normal para $P(X = x)$ que determina la probabilidad de un intervalo de longitud uno (igual al incremento de la variable aleatoria binomial), de manera que el punto medio del intervalo sea igual al valor x. Por lo tanto,

$$P(X_B = x) \approx P\left(\frac{x - np - 1/2}{\sqrt{np(1-p)}} \leq Z_N \leq \frac{x - np + 1/2}{\sqrt{np(1-p)}}\right).$$

Como resultado, la expresión (5.17) puede modificarse de la siguiente forma:

$$P(a \leq X_B \leq b) \approx P\left(\frac{a - np - 0.5}{\sqrt{np(1-p)}} \leq Z_N \leq \frac{b - np + 0.5}{\sqrt{np(1-p)}}\right). \tag{5.18}$$

TABLA 5.4 Frecuencias relativas observada y teórica para la demanda diaria de un producto

Demanda diaria	*Frecuencia relativa*	*Intervalo normal estándar*	*Probabilidad del intervalo*
(55–64)	0.0588	(−3.75– −2.92)	0.0017
(65–74)	0.0392	(−2.92– −2.08)	0.0170
(75–84)	0.0588	(−2.08– −1.25)	0.0868
(85–94)	0.1961	(−1.25– −0.42)	0.2316
(95–104)	0.3137	(−0.42–0.42)	0.3256
(105–114)	0.1765	(0.42–1.25)	0.2316
(115–124)	0.0588	(1.25–2.08)	0.0868
(125–134)	0.0588	(2.08–2.92)	0.0170
(135–144)	0.0392	(2.92–3.75)	0.0017
Totales	0.9999		0.9998

Ejemplo 5.7 Una organización política planea llevar a cabo una encuesta para detectar la preferencia de los votantes con respecto a los candidatos A y B que ocuparán un puesto en la administración pública. Supóngase que toma una muestra aleatoria de mil ciudadanos. ¿Cuál es la probabilidad de que 550 o más de los votantes indiquen una preferencia por el candidato A si la población, con respecto a los candidatos, se encuentra igualmente dividida?

Sea X la variable aleatoria que representa el número de ciudadanos que tienen preferencia por el candidato A. La muestra aleatoria de mil votantes puede pensarse como un conjunto de ensayos independientes con una probabilidad de éxito, en cada ensayo, igual a 0.5 (candidato A), dado que, por hipótesis, la población de votantes se encuentra igualmente dividida entre los candidatos. De esta forma, X es una variable aleatoria binomial con media $np = 500$ y desviación estándar $\sqrt{np(1 - p)} = 15.81$. La probabilidad de que $X \geq 550$ se puede aproximar, de manera adecuada, mediante el empleo de la distribución normal dado que n es suficientemente grande: grande:

$$
\begin{aligned}
P(X \geq 550) &\approx P[Z_N \geq (549.5 - 500)/15.81] \\
&\approx P(Z_N \geq 3.13) \\
&\approx 0.0009.
\end{aligned}
$$

Como la probabilidad de tal hecho es muy pequeña, si p es igual a 0.5 puede concluirse que A será el ganador en la encuesta, ya que 550 o más personas indicarán una preferencia por él.

5.3 La distribución uniforme

Supóngase que ocurre un evento en que una variable aleatoria toma valores de un intervalo finito, de manera que éstos se encuentran distribuidos igualmente sobre el intervalo. Esto es, la probabilidad de que la variable aleatoria tome un valor en cada subintervalo de igual longitud es la misma. Se dice entonces que la variable aleatoria se encuentra *distribuida uniformemente* sobre el intervalo.

Definición 5.2 Se dice que una variable aleatoria X está distribuida uniformemente sobre el intervalo (a, b) si su función de densidad de probabilidad está dada por:

$$
f(x; a, b) = \begin{cases} 1/(b - a) & a \leq x \leq b, \\ 0 & \text{para cualquier otro valor} \end{cases} \tag{5.19}
$$

La función de densidad de probabilidad de una distribución uniforme es constante en el intervalo (a, b), como se ilustra en la figura 5.5. Por esto, tal distribución también se conoce como distribución "rectangular".

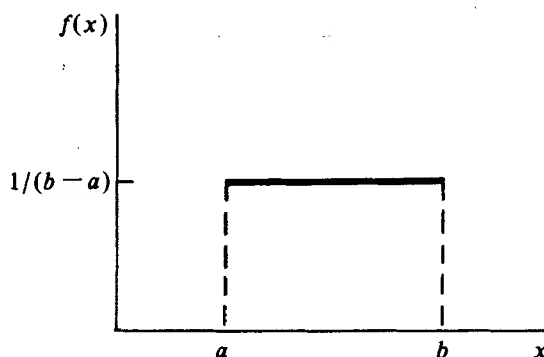


FIGURA 5.5 Gráfica de la función de densidad de probabilidad uniforme

La función de distribución acumulativa se determina de manera fácil y está dada por

$$P(X \leq x) = F(x; a, b) = (b - a)^{-1} \int_a^x dt$$

$$= \begin{cases} 0 & x < a, \\ (x - a)/(b - a) & a \leq x \leq b, \\ 1 & x > b. \end{cases} \qquad (5.20)$$

Se sigue entonces que, para cualquier subintervalo (a_1, b_1) interior a (a, b):

$$P(a_1 \leq X \leq b_1) = F(b_1; a, b) - F(a_1; a, b)$$

$$= (b_1 - a_1)/(b - a). \qquad (5.21)$$

Este resultado ilustra que la probabilidad de que X tome valores del subintervalo (a_1, b_1) es $1/(b - a)$ por la longitud del subintervalo y, de esta forma, igual a la probabilidad de que X tome un valor en cualquier otro subintervalo de la misma longitud.

La distribución uniforme proporciona una representación adecuada para redondear las diferencias que surgen al medir cantidades físicas entre los valores observados y los reales. Por ejemplo, si el peso de un individuo se redondea al kilogramo más cercano, entonces la diferencia entre éste y el peso verdadero será algún valor entre −0.5 y 0.5 kg. Es común que el error de redondeo se encuentra distribuido uniformemente en el intervalo (−0.5, 0.5). Otro uso de la distribución uniforme es proporcionar una aproximación clara sobre un intervalo muy pequeño cuya distribución es distinta a la uniforme.

Ejemplo 5.8 Con respecto al ejemplo 1.1, si se supone que las cuotas se encuentran distribuidas de manera uniforme en el intervalo ($81.5-$111.5), entonces la función

de densidad de probabilidad se determina por:

$$f(x; 81.5, 111.5) = 1/30, \qquad 81.5 \leq x \leq 111.5.$$

Se sigue de (5.21) que la probabilidad de que una cuota se encuentre en un subintervalo de longitud \$5 (la amplitud de clase en el ejemplo 1.1) es 5/30. En la tabla 5.5 se proporciona una comparación entre las frecuencias relativas dadas en la tabla 1.1 y las correspondientes probabilidades teóricas, con base en la distribución uniforme. Como puede observarse, la concordancia entre las frecuencias teóricas y observadas es aparente.

El valor esperado de una variable aleatoria distribuida de manera uniforme es

$$E(X) = (b - a)^{-1} \int_a^b x dx$$

$$= (a + b)/2. \tag{5.22}$$

Para obtener los momentos superiores de X, es más fácil trabajar con la variable aleatoria $Y = X - [(a + b)]/2$, que desplaza la media a cero, dado que $E(Y) = E(X) - [(a + b)]/2$. De esta forma:

$$f(y; \theta) = 1/\theta, \qquad -\theta/2 \leq y \leq \theta/2, \tag{5.23}$$

en donde $\theta = b - a$. De acuerdo con lo anterior, el r-ésimo momento central de Y es igual al r-ésimo momento central alrededor del cero, esto es:

$$\mu_r(Y) = \mu_r'(Y) = \theta^{-1} \int_{-\theta/2}^{\theta/2} y^r dy$$

$$= \left(\frac{1}{\theta}\right) \cdot \frac{y^{r+1}}{r+1}\Bigg|_{-\theta/2}^{\theta/2}$$

$$= \begin{cases} 0 & \text{si } r \text{ es impar} \\ \theta^r/[(r+1)2^r] & \text{si } r \text{ es par.} \end{cases} \tag{5.24}$$

TABLA 5.5 Comparación entre las frecuencias teórica y observada para una distribución uniforme

Cuota anual	*Número observado*	*Frecuencia relativa*	*Intervalo uniforme*	*Probabilidad del intervalo*	*Número esperado*
82– 86	3	0.075	81.5– 86.5	0.167	6.667
87– 91	7	0.175	86.5– 91.5	0.167	6.667
92– 96	8	0.200	91.5– 96.5	0.167	6.667
97–101	8	0.200	96.5–101.5	0.167	6.667
102–106	7	0.175	101.5–106.5	0.167	6.667
107–111	7	0.175	106.5–111.5	0.167	6.667
Totales	40	1.000		1.000	40.000

Dado que ni la varianza ni los factores de forma se ven afectados por el cambio de localización, la varianza, el coeficiente de asimetría y la curtosis relativa de la variable aleatoria distribuida uniformemente se encuentran a partir de **(5.24)** y están determinadas por:

$$Var(X) = (b - a)^2/12, \tag{5.25}$$

$$\alpha_3(X) = 0, \text{ y} \tag{5.26}$$

$$\alpha_4(X) = \frac{(b - a)^4/80}{[(b - a)^2/12]^2} = \frac{9}{5}. \tag{5.27}$$

Puede emplearse (5.23) para determinar la desviación media de la siguiente manera:

$$\begin{aligned} E|Y| &= \theta^{-1} \int_{-\theta/2}^{\theta/2} |y| dy \\ &= 2\theta^{-1} \int_0^{\theta/2} y dy \\ &= \theta/4. \end{aligned} \tag{5.28}$$

De esta forma la desviación media de una variable aleatoria distribuida de manera uniforme está dada por $(b - a)/4$.

Una distribución uniforme es simétrica y tiene un pico menor que el de la distribución normal, no tiene moda y su mediana es igual a la media. Los valores cuantiles x_q, correspondientes a la proporción acumulativa q, son de manera tal que:

$$F(x_q; a, b) = q,$$

los que, por (5.20) son:

$$x_q = a + (b - a)q. \tag{5.29}$$

En la tabla 5.6 se encuentran resumidas las propiedades de esta distribución.

Más adelante se examinará el caso especial cuando $a = 0$ y $b = 1$. Este último se conoce como distribución uniforme sobre el intervalo unitario (0, 1) con función de

TABLA 5.6 Propiedades básicas de la distribución uniforme

Función de densidad de probabilidad	*Parámetros*	
$f(x; a, b) = 1/(b - a), \quad a \le x \le b$	a,	$-\infty < a < \infty$
	b,	$-\infty < b < \infty$.

Media	*Varianza*	*Desviación media*	*Valor del cuantil*	*Coeficiente de asimetría*	*Curtosis relativa*
$(a + b)/2$	$(b - a)^2/12$	$(b - a)/4$	$x_q = a + (b - a)q$	0	9/5

densidad de probabilidad:

$$f(x; 0, 1) = 1, \qquad 0 \leq x \leq 1. \tag{5.30}$$

Esta distribución es, de manera especial, muy importante ya que tiene un papel clave en la simulación por computadora de los valores de una variable aleatoria con una distribución específica.

5.4 La distribución beta

Una distribución que permite generar una gran variedad de perfiles es la *distribución beta*. Se ha utilizado para representar variables físicas cuyos valores se encuentran restringidos a un intervalo de longitud finita y para encontrar ciertas cantidades que se conocen como límites de tolerancia sin necesidad de la hipótesis de una distribución normal. Además, la distribución beta juega un gran papel en la estadística bayesiana. Se examinará un ejemplo de lo anterior en el capítulo seis.

Definición 5.3 Se dice que una variable aleatoria X posee una distribución beta si su función de densidad de probabilidad está dada por:

$$f(x; \alpha, \beta) = \begin{cases} \dfrac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1 - x)^{\beta-1} & 0 < x < 1, \quad \alpha, \beta > 0, \\ 0 & \text{para cualquier otro valor} \end{cases} \tag{5.31}$$

Las cantidades α y β de la distribución beta son, ambas, parámetros de perfil. Valores distintos de α y β darán distintos perfiles para la función de densidad beta. Sin tanto α como β son menores que uno, la distribución beta tiene un perfil en forma de U. Si $\alpha < 1$ y $\beta \geq 1$, la distribución tiene un perfil de J transpuesta, y si $\beta < 1$ y $\alpha \geq 1$, el perfil es una J. Cuando tanto α y β son ambos mayores que uno, la distribución presenta un pico en $x = (\alpha - 1)/(\alpha + \beta - 2)$. Finalmente, la distribución beta es simétrica cuando $\alpha = \beta$. En la figura 5.6 se encuentran ilustrados estos perfiles para valores específicos de α y β. Nótese que si en (5.31) x se reemplaza por $x - 1$, se obtiene la siguiente relación de simetría

$$f(1 - x; \beta, \alpha) = f(x; \alpha, \beta) \tag{5.32}$$

El nombre de esta distribución proviene de su asociación con la función beta que se encuentra definida por

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1 - x)^{\beta-1} dx. \tag{5.33}$$

Puede demostrarse que las funciones beta y gama se encuentran relacionadas por la expresión

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}. \tag{5.34}$$

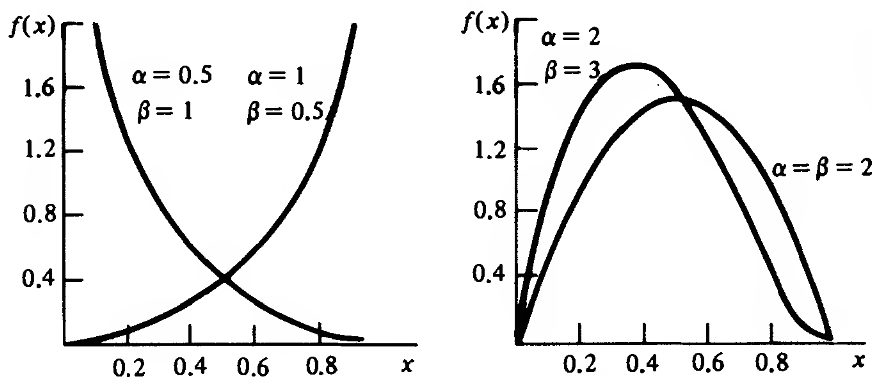


FIGURA 5.6 Gráficas de la función de densidad beta para distintos valores de α y β

Mediante el empleo de (5.33) y (5.34), es obvio que (5.31) es una función de densidad de probabilidad. Esto es:

$$\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha-1}(1 - x)^{\beta-1}dx = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} B(\alpha, \beta) = 1,$$

y puesto que $f(x; \alpha, \beta)$ es no negativa, (5.31) es una función de densidad de probabilidad.

La función de distribución acumulativa se encuentra definida por:

$$P(X \leq x) = F(x; \alpha, \beta) = \begin{cases} 0 & x \leq 0, \\ \dfrac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \displaystyle\int_0^x t^{\alpha-1}(1 - t)^{\beta-1}dt & 0 < x < 1, \\ 1 & x \geq 1. \end{cases} \quad (5.35)$$

La integral que aparece en (5.35) es la función beta incompleta:

$$B_x(\alpha, \beta) = \int_0^x t^{\alpha-1}(1 - t)^{\beta-1}dt. \quad (5.36)$$

De esta forma, la función de distribución beta puede expresarse como un cociente de funciones beta incompletas,

$$\begin{aligned} F(x; \alpha, \beta) &= B_x(\alpha, \beta)/B(\alpha, \beta) \\ &= I_x(\alpha, \beta) \qquad 0 < x < 1, \end{aligned} \quad (5.37)$$

donde $I_x(\alpha, \beta)$ se encuentra tabulada de manera extensa (véase [5,6]). En [5], los valores cuantiles x son aquellos para los que $I_x(\alpha, \beta)$ es igual a 0.0025, 0.005, 0.01,

0.025, 0.05, 0.1, 0.25 y 0.5 para las distintas combinaciones de α y β. Con el fin de encontrar los valores cuantiles correspondientes a puntos de alto porcentaje, considérese lo siguiente:

$$\begin{aligned} P(X \leq x) &= P(1 - X \geq 1 - x) \\ &= 1 - P(1 - X \leq 1 - x); \end{aligned}$$

entonces, por la relación de simetría (5.32):

$$F(x; \alpha, \beta) = 1 - F(1 - x; \beta, \alpha)$$

o

$$I_x(\alpha, \beta) = 1 - I_{1-x}(\beta, \alpha). \tag{5.38}$$

De esta manera, los valores cuantiles para los puntos de alto porcentaje se encuentran al intercambiar α y β y toman el punto de porcentaje igual a $1 - x$. A manera de ilustración, sea X una variable aleatoria beta con $\alpha = 2$ y $\beta = 4$; los valores cuantiles 90, 95 y 99 son 0.58389, 0.65741 y 0.77793, respectivamente. En la tabla 5.7 se proporcionan los valores cuantiles para combinaciones de valores de α y β que dan origen a los distintos perfiles de la distribución beta.

Es más facil obtener los momentos de la variable aleatoria beta mediante el empleo del método directo, que por el uso de la función generadora de momentos, debido a que esta última no tiene una forma sencilla. En particular, se encontrará una expresión general que permita obtener el r-ésimo momento alrededor del cero y después emplearla para obtener los momentos restantes:

$$\begin{aligned} \mu_r' = E(X^r) &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha + r - 1}(1 - x)^{\beta - 1} dx \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} B(\alpha + r, \beta) \\ &= \frac{\Gamma(\alpha + b)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha + r)\Gamma(\beta)}{\Gamma(\alpha + \beta + r)} \\ &= \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + r)}{\Gamma(\alpha)\Gamma(\alpha + \beta + r)}. \end{aligned} \tag{5.39}$$

Como resultado,

$$\begin{aligned} E(X) &= \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + 1)}{\Gamma(\alpha)\Gamma(\alpha + \beta + 1)} \\ &= \frac{\alpha}{\alpha + \beta}, \end{aligned} \tag{5.40}$$

y

$$\begin{aligned} Var(X) &= \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)} - \frac{\alpha^2}{(\alpha + \beta)^2} \\ &= \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}. \end{aligned} \tag{5.41}$$

TABLA 5.7 Valores de cuantiles beta para distintas combinaciones de α y β

	$x_{0.25}$	$x_{0.50}$	$x_{0.75}$
$\alpha = \beta = 1/2$	0.14645	0.50000	0.85355
$\alpha = 1/2, \beta = 2$	0.02831	0.12061	0.31122
$\alpha = 2, \beta = 1/2$	0.68878	0.87939	0.97169
$\alpha = 4, \beta = 6$	0.29099	0.39308	0.50199

Al seguir este procedimiento y después de efectuar el álgebra necesaria, el coeficiente de asimetría y la curtosis relativa para la distribución beta están dadas por:

$$\alpha_3(X) = \frac{2(\beta - \alpha)\sqrt{\alpha + \beta + 1}}{\sqrt{\alpha\beta}\,(\alpha + \beta + 2)}, \tag{5.42}$$

y

$$\alpha_4(X) = \frac{3(\alpha + \beta + 1)[2(\alpha + \beta)^2 + \alpha\beta(\alpha + \beta - 6)]}{\alpha\beta(\alpha + \beta + 2)(\alpha + \beta + 3)}. \tag{5.43}$$

Mediante el examen de (5.42) puede observarse que la distribución beta es simétrica sólo si $\alpha = \beta$, tal y como ya se había mencionado. Si $\alpha < \beta$, la distribución tiene un sesgo positivo y si $\alpha > \beta$, la distribución presenta un sesgo negativo.

En la tabla 5.8 se proporciona un resumen de las propiedades de la distribución beta.

Algunas áreas, en las que se emplea la distribución beta como modelo de probabilidad incluyen la distribución de artículos defectuosos sobre un intervalo de tiempo específico; la distribución del intervalo de tiempo necesario para completar una fase de proyecto en PERT, evaluación de programas y técnicas de revisión, (en este caso se emplea la distribución beta generalizada; véase [14]); la distribución de la proporción de los valores que deben caer entre dos observaciones extremas.

TABLA 5.8 Propiedades básicas de la distribución beta

Función de densidad de probabilidad		*Parámetros*	
$f(x; \alpha, \beta) = \dfrac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1 - x)^{\beta-1}$, $0 < x < 1$		α, $\alpha > 0$; β, $\beta > 0$	
Media	*Varianza*	*Coeficiente de asimetría*	*Curtosis relativa*
$\dfrac{\alpha}{\alpha + \beta}$	$\dfrac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$	$\dfrac{2(\beta - \alpha)\sqrt{\alpha + \beta + 1}}{\sqrt{\alpha\beta}\,(\alpha + \beta + 2)}$	*

$*\dfrac{3(\alpha + \beta + 1)[2(\alpha + \beta)^2 + \alpha\beta(\alpha + \beta - 6)]}{\alpha\beta(\alpha + \beta + 2)(\alpha + \beta + 3)}$

La esencia de esta última área tiene relación con los límites estadísticos de tolerancia. Estos límites son muy importantes, especialmente en el control estadístico de calidad donde el control de variabilidad de un producto es esencial. Este control, en general, se lleva a cabo mediante la medición de algunas propiedades del producto o determinando los ajustes que deben hacerse al proceso de producción para mejorar la calidad del producto. Los límites estadísticos de tolerancia no son iguales a las tolerancias físicas o especificaciones límite. Éstos son conjuntos de criterios diseñados para un proceso de producción en particular y que se espera que todas las unidades cumplan. Los límites estadísticos de tolerancia se tratarán en el capítulo ocho.

Puede demostrarse que si la suma de los parámetros que determinan el perfil de la distribución beta es, de manera relativa, grande, la función de distribución acumulativa beta (5.35) se puede aproximar de manera adecuada por la diferencia de dos funciones de distribución normal estándar. Esto es:

$$F(x; \alpha, \beta) \approx F_N(z_u; 0, 1) - F_N(z_\ell; 0, 1), \tag{5.44}$$

en donde:

$$z_u = \frac{[\beta] - 0.5 - (\alpha + \beta - 1)(1 - x)}{[(\alpha + \beta - 1)(x)(1 - x)]^{1/2}},$$

$$z_\ell = -\frac{(\alpha + \beta - 1)(1 - x) + 0.5}{[(\alpha + \beta - 1)(x)(1 - x)]^{1/2}},$$

y $[\beta]$ denota el entero más grande que no excede a β. En la tabla 5.9 se tiene una comparación entre los valores de la función beta dados por (5.35) con aquéllos proporcionados por (5.44). Para cada valor x, el primer renglón correspondiente a ésta es el valor exacto de la distribución beta y el siguiente es el que proporciona (5.44). Para valores distintos de los finales, la aproximación es adecuada. Sin embargo, nótese que la discrepancia en los valores superiores disminuye conforme la suma de α y β es más grande.

TABLA 5.9 Comparación entre las funciones de distribución beta y normal

x	$\alpha = \beta = 5$	$\alpha = 10, \beta = 5$	$\alpha = 10, \beta = 15$
0.10	0.0008909	0.0000001	0.0000521
	0.0000317	0.0	0.0000007
0.25	0.04893	0.0003419	0.05466
	0.04182	0.0001078	0.04947
0.50	0.50	0.08978	0.8463
	0.4996	0.09009	0.8461
0.75	0.95107	0.74153	0.99989
	0.94118	0.72564	0.99886
0.90	0.9991091	0.99077	1.0
	0.9405883	0.95160	0.9756

5.5 La distribución gama

Otra distribución de gran uso es la *distribución gama*. Entre los muchos usos que esta distribución tiene se encuentra el siguiente: supóngase que una pieza metálica se encuentra sometida a cierta fuerza, de manera que se romperá después de aplicar un número específico de ciclos de fuerza. Si los ciclos ocurren de manera independiente y a una frecuencia promedio, entonces el tiempo que debe transcurrir antes de que el material se rompa es una variable aleatoria que cumple con la distribución gama.

Definición 5.4 Se dice que la variable aleatoria X tiene una distribución gama si su función de densidad de probabilidad está dada por:

$$f(x;\alpha,\theta)=\begin{cases}\dfrac{1}{\Gamma(\alpha)\theta^{\alpha}}x^{\alpha-1}\exp(-x/\theta) & x>0,\quad \alpha,\theta>0\\ 0 & \text{para cualquier otro valor,}\end{cases} \tag{5.45}$$

en dond^ $\Gamma(\alpha)$ es la función gama definida en el capítulo tres.

La distribución gama es muy versátil puesto que exhibe varios perfiles que dependen del valor del parámetro α. En la figura 5.7 se ilustran distintos perfiles de la función de densidad gama para distintos valores de α y θ. Como puede observarse, para $\alpha \leq 1$, la distribución gama tiene un perfil en forma de J transpuesta. Para

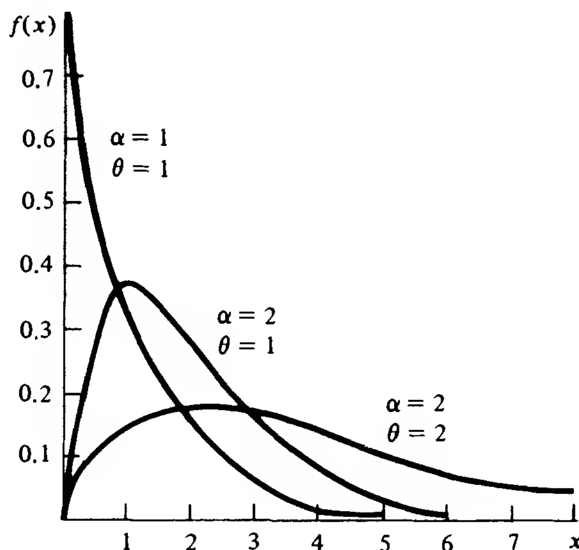


FIGURA 5.7 Gráficas de la funcion de densidad gama para distintos valores de α y θ

$\alpha > 1$, presenta un pico que ocurre en $x = \theta(\alpha - 1)$. Para un valor fijo de θ, el perfil básico de la distribución gama no se altera si el valor de α cambia. Lo anterior da como resultado que las cantidades α y θ son los factores de forma y de escala, respectivamente, de la distribución gama.

Esta distribución se emplea de manera extensa en una gran diversidad de áreas; por ejemplo, para representar el tiempo aleatorio de falla de un sistema que falla sólo si de manera exacta los componentes fallan y la falla de cada componente ocurre a una frecuencia constante $\lambda = 1/\theta$ por unidad de tiempo. También se emplea en problemas de líneas de espera para representar el intervalo total para completar una reparación si ésta se lleva a cabo en subestaciones; completar la reparación en cada subestación es un evento independiente que ocurre a una frecuencia constante igual a $\lambda = 1/\theta$. Existen algunos ejemplos que no siguen el patrón anterior, pero que se aproximan de manera adecuada mediante el empleo de la distribución gama, como los ingresos familiares y la edad del hombre al contraer matrimonio por primera vez.

Mediante el empleo de la función gama dada por (3.5), puede demostrarse que (5.45) es una función de densidad de probabilidad. Para hacerlo, considérese un cambio de variable de integración, tal que $u = x/\theta$, $x = \theta u$, y $dx = \theta du$; entonces:

$$\frac{1}{\Gamma(\alpha)\theta^{\alpha}}\int_0^{\infty} x^{\alpha-1}\exp(-x/\theta)dx = \frac{1}{\Gamma(\alpha)\theta^{\alpha}}\int_0^{\infty}(\theta u)^{\alpha-1}\exp(-u)\theta du$$
$$= \frac{1}{\Gamma(\alpha)}\int_0^{\infty} u^{\alpha-1}\exp(-u)du = 1,$$

dado que $\Gamma(\alpha) = \int_0^{\infty} u^{\alpha-1}\exp(-u)du$.

Con un procedimiento similar se demuestra que el r-ésimo momento alrededor del cero es:

$$\mu_r' = \frac{1}{\Gamma(\alpha)\theta^{\alpha}}\int_0^{\infty} x^{\alpha+r-1}\exp(-x/\theta)dx$$
$$= \frac{\theta^{\alpha+r}}{\Gamma(\alpha)\theta^{\alpha}}\int_0^{\infty} u^{\alpha+r-1}\exp(-u)du$$
$$= \frac{\theta^{r}\Gamma(\alpha + r)}{\Gamma(\alpha)} \tag{5.46}$$

Se sigue, por lo tanto, que:

$$E(X) = \alpha\theta \tag{5.47}$$

y

$$Var(X) = \alpha\theta^2 \tag{5.48}$$

Además, después de obtener los momentos centrales apropiados, se puede demostrar que el coeficiente de asimetría es

$$\alpha_3(X) = 2/\sqrt{\alpha}. \tag{5.49}$$

y la curtosis relativa está dada por:

$$\alpha_4(X) = 3\left(1 + \frac{2}{\alpha}\right). \tag{5.50}$$

Nótese que a partir de los factores de forma $\alpha_3(X)$ y $\alpha_4(X)$, la distribución gama tiene un sesgo positivo y más picos que la distribución normal, puesto que $\alpha_4(X) > 3$ para cualquier $\alpha > 0$. Sin embargo, también debe notarse que conforme el parámetro α se hace cada vez más grande, el sesgo se convierte en menos pronunciado y la curtosis relativa tiene el tres como valor límite. De hecho, para valores grandes de α la distribución gama puede aproximarse, en algún grado, por una distribución normal. Esto es, la variable aleatoria

$$Z = (X - \alpha\theta)/\theta\sqrt{\alpha} \tag{5.51}$$

es, de manera aproximada, igual a la normal estándar para valores grandes de α.

La función generadora de momentos para la variable aleatoria gama X está dada por:

$$E[\exp(tX)] = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} \int_0^{\infty} x^{\alpha-1}\exp[-(1 - \theta t)x/\theta]dx.$$

Sea $u = (1 - \theta t)x/\theta$, $x = u\theta/(1 - \theta t)$, y $dx = [\theta/(1 - \theta t)]du$. Entonces:

$$E[\exp(tX)] = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} \int_0^{\infty} \frac{u^{\alpha-1}\theta^{\alpha-1}}{(1 - \theta t)^{\alpha-1}} \exp(-u) \frac{\theta}{(1 - \theta t)} du$$

$$= \frac{1}{\Gamma(\alpha)(1 - \theta t)^{\alpha}} \int_0^{\infty} u^{\alpha-1}\exp(-u)du$$

$$= (1 - \theta t)^{-\alpha}, \qquad 0 \leq t < 1/\theta. \tag{5.52}$$

La función de distribución acumulativa se determina por la siguiente expresión:

$$F(x; \alpha, \theta) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} \int_0^{x} t^{\alpha-1}\exp(-t/\theta)dt, \qquad x > 0. \tag{5.53}$$

Se tabularon muchas versiones de (5.53). Por ejemplo, si se efectúa el cambio de variable $u = t/\theta$ de manera tal que $t = \theta u$ y $dt = \theta du$, entonces (5.53) toma la siguiente forma:

$$F(x; \alpha, \theta) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} \int_0^{x/\theta} (\theta u)^{\alpha-1}\exp(-u)\theta du$$

$$= \frac{1}{\Gamma(\alpha)} \int_0^{x/\theta} u^{\alpha-1}\exp(-u)du.$$

La integral $\int_0^{x/\theta} u^{\alpha-1}\exp(-u)du$ se conoce como la *función gama incompleta* y se denota, generalmente, por $\gamma(x/\theta; \alpha)$. El cociente de $\gamma(x/\theta; \alpha)$ y de la función gama completa $\Gamma(\alpha)$ recibe el nombre de *cociente de la función gama incompleta* y

se encuentra tabulado en [8] para distintos valores de x/θ y α. De acuerdo con lo anterior, la función gama de distribución acumulativa se escribe como:

$$P(X \leq x) = F(x; \alpha, \theta) \doteq \gamma(x/\theta; \alpha)/\Gamma(\alpha). \tag{5.54}$$

En [7] se encuentra una tabla muy extensa de los valores de una función equivalente a (5.53), dada por:

$$I(u, p) = F(x; \alpha, \theta), \tag{5.55}$$

donde $u = x/\theta\sqrt{\alpha}$ y $p = \alpha - 1$. Debe notarse que si el parámetro de forma α es un entero positivo, (5.55) se puede expresar, en forma cerrada:

$$F(x; \alpha, \theta) = 1 - \left[1 + \frac{x}{\theta} + \frac{1}{2!}\left(\frac{x}{\theta}\right)^2 + \cdots + \frac{1}{(\alpha - 1)!}\left(\frac{x}{\theta}\right)^{\alpha - 1}\right]\exp(-x/\theta) \tag{5.56}$$

como resultado de efectuar varias integraciones por partes. También el valor cuantil x_q para el que $F(x_q; \alpha, \theta) = q$ no puede determinarse de manera directa; éste puede interpolarse a partir de los valores que aparecen en las tablas dadas en [7] μ [8]. En la tabla 5.10 se da un breve resumen de las propiedades básicas de la distribución gama.

Ejemplo 5.9 Supóngase que cierta pieza metálica se romperá después de sufrir dos ciclos de esfuerzo. Si estos ciclos ocurren de manera independiente a una frecuencia promedio de dos por cada 100 horas, obtener la probabilidad de que el intervalo de tiempo se encuentre hasta que ocurre el segundo ciclo: *a)* dentro de una desviación estándar del tiempo promedio, y *b)* a más de dos desviaciones estándar por encima de la media.

Sea X la variable aleatoria que representa el lapso que transcurre hasta que la pieza sufre el segundo ciclo de esfuerzo. Si X tiene una distribución gama con $\alpha = 2$ y $\theta = 50$ horas debido a que la frecuencia promedio es 0.02 por hora. La fun-

TABLA 5.10 Propiedades de la distribución gama

Función de densidad de probabilidad	*Parámetros*
$f(x; \alpha, \theta) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1}\exp(-x/\theta)$	α, $\alpha > 0$
$x > 0$	θ, $\theta > 0$

Media	*Varianza*	*Coeficiente de asimetría*	*Curtosis relativa*
$\alpha\theta$	$\alpha\theta^2$	$2/\sqrt{\alpha}$	$3\left(1 + \frac{2}{\alpha}\right)$

ción de densidad de probabilidad es

$$f(x; 2, 50) = \frac{1}{\Gamma(2)50^2} x \exp(-x/50), \qquad x > 0,$$

y la función de distribución acumulativa dada por (5.56) se reduce a:

$$F(x; \alpha, \theta) = 1 - \left(1 + \frac{x}{50}\right)\exp(-x/50), \qquad x > 0.$$

De (5.47) y (5.48), los valores de la media y de la desviación estándar de X son 100 y 70.71, respectivamente. De acuerdo con lo anterior:

$$\begin{aligned} P(\mu - \sigma < X < \mu + \sigma) &= P(29.29 < X < 170.71) \\ &= F(170.71; 2, 50) - F(29.29; 2, 50) \\ &= 0.7376. \end{aligned}$$

Nótese que la probabilidad de que el lapso sea menor de una desviación estándar por debajo de la media es de 0.1172 y la probabilidad de que éste sea más grande que la media por una desviación estándar es $1 - 0.8548 = 0.1452$. Finalmente:

$$\begin{aligned} P(X > \mu + 2\sigma) &= P(X > 241.42) \\ &= 1 - F(241.42; 2, 50) \\ &= 0.0466. \end{aligned}$$

Ejemplo 5.10 Para demostrar el grado de concordancia entre las distribuciones normal y gama, se seleccionaron, para esta última, los valores de 3.5 y 7 para el parámetro de forma α, y para $\theta = 10$, calculándose las funciones de distribución acumulativa para distintos valores de las correspondientes variables aleatorias. La información anterior se encuentra en la tabla 5.11.

A partir de la información dada en la tabla 5.11, es evidente que la función de distribución acumulativa normal sobreestima los valores dados por la correspondiente función de distribución acumulativa gama en los extremos, mientras que la subestima alrededor de la media. Lo anterior es válido para los dos valores de α; sin embargo, para $\alpha = 7$, la concordancia en los extremos es considerablemente mejor que cuando $\alpha = 3.5$. Como resultado, se espera que la concordancia aumente para valores de α más grandes que siete.

Cuando α es un entero positivo, la distribución gama también se conoce como *distribución de Erlang* en honor del científico danés que la usó por primera vez a principios del año 1900 a fin de establecer resultados útiles para problemas de tráfico en líneas telefónicas. Existe una asociación entre los modelos de probabilidad de Poisson y de Erlang. Si el número de eventos aleatorios independientes que ocurren en un lapso específico es una variable de Poisson con una frecuencia constante de ocurrencia igual a $1/\theta$, entonces para una α, el tiempo de espera hasta que ocurre el α-ésimo evento de Poisson tiene una distribución de Erlang. Este resultado se sigue al hacer una comparación entre las funciones de distribución acumulativa de los mo-

TABLA 5.11 Comparación entre las funciones de distribución acumulativa gama y normal

$\alpha = 3.5, \theta = 10, p = 2.5; \mu = 35, \sigma = 18.71$				$\alpha = 7, \theta = 10, p = 6; \mu = 70, \sigma = 26.46$			
X	u	*Gama* $I(u, p)$	*Normal* $F(x; \mu, \sigma)$	X	u	*Gama* $I(u, p)$	*Normal* $F(x; \mu, \sigma)$
0	0	0	0.0307	0	0	0	0.0041
5	0.27	0.0058	0.0516	10	0.38	0.000098	0.0116
10	0.53	0.0397	0.0902	20	0.76	0.004865	0.0294
15	0.80	0.1144	0.1423	30	1.13	0.0431	0.0655
20	1.07	0.2209	0.2119	40	1.51	0.1103	0.1292
25	1.34	0.3417	0.2981	50	1.89	0.2380	0.2236
30	1.60	0.4587	0.3936	60	2.27	0.3946	0.3520
35	1.87	0.5706	0.5000	70	2.65	0.5518	0.5000
40	2.14	0.6678	0.6064	80	3.02	0.6853	0.6480
45	2.41	0.7485	0.7019	90	3.40	0.7928	0.7764
50	2.67	0.8107	0.7881	100	3.78	0.8698	0.8708
55	2.94	0.8612	0.8577	110	4.16	0.9215	0.9345
60	3.21	0.8997	0.9098	120	4.54	0.9544	0.9706
65	3.47	0.9274	0.9485	130	4.91	0.9739	0.9884
70	3.74	0.9486	0.9693	140	5.29	0.9857	0.9959
75	4.01	0.9640	0.9838	150	5.67	0.9924	0.9987
80	4.28	0.9750	0.9920	160	6.05	0.9960	0.9997

delos de Poisson y de Erlang dadas por (4.17) y (5.56), respectivamente. Esto es, la probabilidad de que ocurran a lo más $\alpha - 1$ eventos de Poisson en un tiempo x a una frecuencia constante $1/\theta$ se desprende de (4.17) y está dado por:

$$F_P(\alpha - 1; x/\theta) = \left[1 + \frac{x}{\theta} + \frac{1}{2!}\left(\frac{x}{\theta}\right)^2 + \cdots + \frac{1}{(\alpha - 1)!}\left(\frac{x}{\theta}\right)^{\alpha - 1}\right]\exp(-x/\theta).$$

Por otro lado, si se supone que el tiempo de espera sigue el modelo de Erlang, la probabilidad de que el tiempo de espera hasta que ocurra el α-ésimo evento exceda un lapso x específico, está determinado por:

$$\begin{aligned} P(X > x) &= 1 - F_L(x; \alpha, \theta) \\ &= 1 - \left\{1 - \left[1 + \frac{x}{\theta} + \frac{1}{2!}\left(\frac{x}{\theta}\right)^2 + \cdots \right.\right. \\ &\qquad \left.\left. + \frac{1}{(\alpha - 1)!}\left(\frac{x}{\theta}\right)^{\alpha - 1}\right]\exp(-x/\theta)\right\} \\ &= \left[1 + \frac{x}{\theta} + \frac{1}{2!}\left(\frac{x}{\theta}\right)^2 + \cdots + \frac{1}{(\alpha - 1)!}\left(\frac{x}{\theta}\right)^{\alpha - 1}\right]\exp(-x/\theta) \\ &= F_P(\alpha - 1; x/\theta). \end{aligned} \tag{5.57}$$

En otras palabras, la probabilidad de que el tiempo que transcurre hasta el α-ésimo evento exceda el valor x es igual a la probabilidad de que el número de eventos de Poisson observados en x no sea mayor que $\alpha - 1$. De esta forma, la distribución de Erlang es el modelo para el tiempo de espera hasta que ocurre el α-ésimo evento de Poisson, y la distribución de Poisson es el modelo para el número de eventos independientes que ocurren en un tiempo x, encontrándose éste distribuido de acuerdo con el modelo de Erlang. En este contexto, $1/\theta$ es la frecuencia constante de ocurrencia y θ es el tiempo promedio entre dos ocurrencias sucesivas.

Cuando el parámetro de forma α es igual a uno, la distribución de Erlang (gama) se reduce a lo que se conoce como la *distribución exponencial negativa.* Esta distribución se emplea de manera extensa para representar lapsos aleatorios de tiempo y se trata con gran detalle en una sección subsecuente de este capítulo. Sin embargo, nótese que la variable aleatoria de una distribución exponencial negativa puede pensarse como el lapso que transcurre hasta el primer evento de Poisson. De acuerdo con lo anterior, la variable aleatoria de Erlang es la suma de variables aleatorias independientes distribuidas exponencialmente.

Otro caso especial del modelo de probabilidad gama es la distribución chi-cuadrado. Si se reemplaza en (5.45) el parámetro de forma α con $\nu/2$ y el parámetro de escala θ con 2, el resultado es la función de densidad de probabilidad de una variable aleatoria chi-cuadrado y se determina por:

$$f(x;\nu) = \begin{cases} \dfrac{1}{\Gamma(\nu/2)2^{\nu/2}} x^{\nu/2-1}\exp(-x/2) & x > 0, \\ 0 & \text{para cualquier otro valor.} \end{cases} \tag{5.58}$$

La distribución chi-cuadrado se encuentra caracterizada por un solo parámetro ν, que recibe el nombre de *grados de libertad.* Como se verá, esta distribución interviene en la inferencia estadística y de manera especial al hacer inferencias con respecto a las varianzas. Se emplea, de manera general, la notación $X \sim \chi^2_\nu$ para indicar que una variable aleatoria tiene una distribución chi-cuadrado con ν grados de libertad.

La función de distribución acumulativa está dada por:

$$P(X \leq x) = \frac{1}{\Gamma(\nu/2)2^{\nu/2}} \int_0^x t^{\nu/2-1} \exp(-t/2)dt \qquad x > 0, \tag{5.59}$$

y se encuentra tabulada de manera extensa. En la tabla E del apéndice se encuentran los valores cuantiles $x_{1-\alpha,\nu}$, de manera que

$$P(X \leq x_{1-\alpha,\nu}) = \int_0^{x_{1-\alpha,\nu}} f(x;\nu)dx = 1 - \alpha$$

para algunas proporciones acumulativas seleccionadas $1 - \alpha$* y distintos valores de ν. A manera de ilustración, si $\nu = 10$,

* En este contexto, la introducción de la cantidad α, $0 \leq \alpha \leq 1$, sirve para facilitar una discusión posterior de un concepto que recibe el nombre de "probabilidad del error de tipo I", que de manera general se denota por α.

$$P(X \leq x_{0.01,10}) = P(X \leq 2.55) = 0.01,$$
$$P(X \leq x_{0.05,10}) = P(X \leq 3.94) = 0.05,$$
$$P(X \leq x_{0.95,10}) = P(X \leq 18.31) = 0.95,$$
$$P(X \leq x_{0.99,10}) = P(X \leq 23.19) = 0.99.$$

Los momentos de la distribución chi-cuadrado se obtienen a partir de (5.47) a (5.50) y están dados por:

$$E(X) = \nu,$$
$$Var(X) = 2\nu,$$
$$\alpha_3(X) = 4/\sqrt{2\nu},$$
$$\alpha_4(X) = 3\left(1 + \frac{4}{\nu}\right).$$

Análogamente y a partir de (5.52), la función generadora de momentos para la distribución chi-cuadrado es:

$$m_X(t) = (1 - 2t)^{-\nu/2} \qquad 0 \leq t < \frac{1}{2}. \tag{5.60}$$

Nótese que una característica interesante de la distribución chi-cuadrado es que el valor de su varianza es dos veces el valor de su media. Además, como está distribución es un caso especial de la distribución gama, presenta un sesgo positivo y un pico mayor que el de una distribución normal, pero el coeficiente de asimetría tiende a cero y a una curtosis relativa igual a tres conforme ν tiende al infinito.

5.6 La distribución de Weibull

La distribución de Weibull fue establecida por el físico suizo del mismo nombre, quien demostró, con base en una evidencia empírica, que el esfuerzo al que se someten los materiales puede modelarse de manera adecuada mediante el empleo de esta distribución [9]. En los últimos 25 años esta distribución se empleó como modelo para situaciones del tipo tiempo-falla y con el objetivo de lograr una amplia variedad de componentes mecánicos y eléctricos.

Definición 5.5 Se dice que una variable aleatoria X tiene una *distribución de Weibull* si su función de densidad de probabilidad está dada por:

$$f(x; \alpha, \theta) = \begin{cases} \dfrac{\alpha}{\theta^\alpha} x^{\alpha-1} \exp[-(x/\theta)^\alpha] & x > 0; \quad \alpha, \theta > 0, \\ 0 & \text{para cualquier otro valor.} \end{cases} \tag{5.61}$$

La distribución de Weibull es una familia de distribuciones que dependen de dos parámetros: el de forma α y el de escala θ. Se puede introducir un parámetro adi-

cional al reemplazar la variable aleatoria de Weibull X por $X - a$, en donde a es un parámetro de localización que representa un valor umbral o tiempo de garantía. En la figura 5.8 se muestran varias gráficas de la distribución de Weibull para distintos valores de α y θ, y como puede observarse, esta distribución tiene distintos perfiles dependiendo del valor de α. Por ejemplo, si $\alpha < 1$, (5.61) tiene una forma de J transpuesta, y si $\alpha > 1$, la función de densidad de Weibull presenta un pico único.

La función de distribución acumulativa de Weibull

$$F(x; \alpha, \theta) = \frac{\alpha}{\theta^\alpha} \int_0^x t^{\alpha-1} \exp[-(t/\theta)^\alpha] dt \tag{5.62}$$

puede obtenerse en forma cerrada mediante la evaluación directa de la integral en (5.62). Esto es:

$$\begin{aligned} F(x; \alpha, \theta) &= \frac{\alpha}{\theta^\alpha}\left(-\frac{\theta^\alpha}{\alpha}\right)\exp[-(t/\theta)^\alpha]\Big|_0^x \\ &= 1 - \exp[-(x/\theta)^\alpha], \qquad x \geq 0. \end{aligned} \tag{5.63}$$

De (5.63), el valor cuantil x_q es:

$$\begin{aligned} 1 - \exp[-(x_q/\theta)^\alpha] &= q \\ x_q &= -\theta[\ln(1 - q)]^{1/\alpha} \\ &= \theta\left[\ln\left(\frac{1}{1-q}\right)\right]^{1/\alpha}. \end{aligned} \tag{5.64}$$

En particular, la mediana de una variable aleatoria de Weibull es:

$$x_{0.5} = \theta[\ln(2)]^{1/\alpha}. \tag{5.65}$$

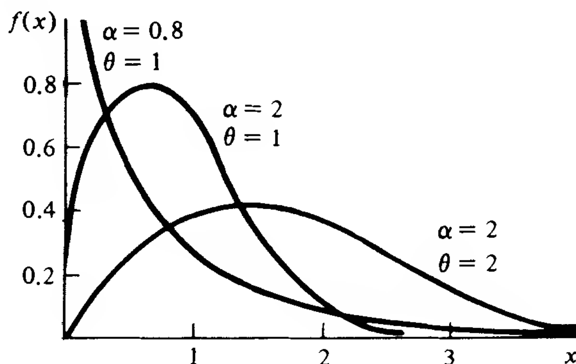


FIGURA 5.8 Gráficas de la función de densidad de Weibull para distintos valores de α y θ

Los momentos y los factores de una variable aleatoria de Weibull se encuentran primero al determinar el r-ésimo momento central alrededor del cero:

$$\mu_r' = E(X^r) = \int_0^\infty x^r f(x; \alpha, \theta)dx$$

$$= \frac{\alpha}{\theta^\alpha}\int_0^\infty x^{\alpha+r-1}\exp[-(x/\theta)^\alpha]dx. \qquad (5.66)$$

En (5.66), sea $u = (x/\theta)^\alpha$; entonces $x = \theta u^{1/\alpha}$ y $dx = (\theta/\alpha)u^{1/\alpha - 1}\,du$. El resultado es:

$$\mu_r' = \frac{\alpha}{\theta^\alpha}\int_0^\infty (\theta u^{1/\alpha})^{\alpha+r-1}\exp(-u)\frac{\theta}{\alpha}u^{1/\alpha-1}du$$

$$= \theta^r\int_0^\infty u^{r/\alpha}\exp(-u)du$$

$$= \theta^r\Gamma\left(1 + \frac{r}{\alpha}\right). \qquad (5.67)$$

De (5.67), la media de X es:

$$E(X) = \theta\Gamma\left(1 + \frac{1}{\alpha}\right), \qquad (5.68)$$

y la varianza de X es el resultado de evaluar:

$$Var(X) = \theta^2\left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right]. \qquad (5.69)$$

Mediante el empleo del mismo procedimiento pueden determinarse el coeficiente de asimetría y la curtosis relativa. Éstos se encuentran en la tabla 5.12. Los facto-

TABLA 5.12 Propiedades básicas de la distribución de Weibull

Función de densidad de probabilidad			*Parámetros*	
$f(x; \alpha, \theta) = \frac{\alpha}{\theta^\alpha}x^{\alpha-1}\exp[-(x/\theta)^\alpha]$			$\alpha, \quad \alpha > 0$	
$x > 0$			$\theta, \quad \theta > 0$	
Media	*Varianza*	*Valor del cuantil*	*Coeficiente de asimetría*	*Curtosis*
$\theta\Gamma\left(1 + \frac{1}{\alpha}\right)$	$\theta^2\left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right]$	$x_q = \theta\left[\ln\left(\frac{1}{1-q}\right)\right]^{1/\alpha}$	*	**

$${}^*\alpha_3(X) = \frac{\Gamma(1 + 3/\alpha) - 3\Gamma(1 + 1/\alpha)\Gamma(1 + 2/\alpha) + 2\Gamma^3(1 + 1/\alpha)}{[\Gamma(1 + 2/\alpha) - \Gamma^2(1 + 1/\alpha)]^{3/2}}$$

$${}^{**}\alpha_4(X) = \frac{\Gamma(1 + 4/\alpha) - 4\Gamma(1 + 1/\alpha)\Gamma(1 + 3/\alpha)}{[\Gamma(1 + 2/\alpha) - \Gamma^2(1 + 1/\alpha)]^2} + \frac{6\Gamma^2(1 + 1/\alpha)\Gamma(1 + 2/\alpha) - 3\Gamma^4(1 + 1/\alpha)}{[\Gamma(1 + 2/\alpha) - \Gamma^2(1 + 1/\alpha)]^2}$$

TABLA 5.13 Comparación entre las funciones de distribución acumulativa de Weibull y normal

	$\alpha = 2.25;\ \theta = 10$		$\alpha = 3.6;\ \theta = 10$		$\alpha = 5.83;\ \theta = 10$	
X	*Weibull*	*Normal* (8.858, 4.128)*	*Weibull*	*Normal* (9.01, 2.788)*	*Weibull*	*Normal* (9.267, 1.828)*
0	0	0.01578	0	0.000619	0	0
1	0.005608	0.02872	0.000251	0.002052	0.000001	0.000003
2	0.026395	0.04746	0.003041	0.006037	0.000084	0.000034
3	0.0644	0.0778	0.013025	0.01539	0.000894	0.000302
4	0.1195	0.1190	0.036259	0.03593	0.004775	0.001988
5	0.1896	0.1762	0.0792	0.07493	0.017425	0.009903
6	0.2716	0.2420	0.1470	0.1401	0.049616	0.03673
7	0.3612	0.3264	0.2419	0.2358	0.1175	0.1075
8	0.4541	0.4150	0.3610	0.3594	0.2384	0.2451
9	0.5457	0.4880	0.4956	0.5000	0.4179	0.4404
10	0.6321	0.6064	0.6321	0.6368	0.6321	0.6554
11	0.7104	0.6985	0.7557	0.7611	0.8250	0.8289
12	0.7785	0.7747	0.8545	0.8599	0.9447	0.9332
13	0.8355	0.8413	0.9236	0.9236	0.9901	0.9793
14	0.8814	0.8925	0.9652	0.9641	0.999184	0.9952
15	0.9171	0.9319	0.9865	0.9842	0.999976	0.999155

* Media y desviación estándar

res de forma pueden graficarse como funciones del parámetro de forma de la distribución de Weibull (véase [2]). Estas gráficas revelan lo siguiente: la distribución de Weibull es simétrica sólo si $\alpha = 3.6$; si $\alpha > 3.6$, la distribución tiene un sesgo negativo y si $\alpha < 3.6$, se encuentra sesgada positivamente. La curtosis relativa se encuentra cercana a la de la distribución normal que es de tres cuando α tiene un valor cercano a 2.25 o a 5.83. En la tabla 5.13 se proporciona una comparación entre las funciones de distribución acumulativa de Weibull y normal, con un α correspondiente a la distribución de 2.25, 3.6 y 5.83 y con un factor de escala $\theta = 10$. La concordancia parece ser relativamente buena tanto en los valores extremos como en el centro, especialmente para $\alpha = 3.6$ y 5.83. De esta forma, la distribución de Weibull puede aproximarse, de manera adecuada, por una distribución normal cada vez que el factor de forma α se encuentre cercano a estos valores.

En la tabla 5.12 se encuentran resumidas propiedades de la distribución de Weibull.

Existen dos casos especiales en la distribución de Weibull que merecen mención especial. Cuando el parámetro de forma es igual a uno, la distribución de Weibull (al igual que la gama), se reduce a la distribución exponencial negativa. Cuando $\alpha = 2$ y el parámetro de escala θ se reemplaza por $\sqrt{2}\,\sigma$, la función de densidad de Weibull (5.61) se reduce a:

$$f(x; \sigma^2) = \frac{x}{\sigma^2} \exp(-x^2/2\sigma^2) \qquad x > 0, \tag{5.70}$$

que es la función de densidad de probabilidad de lo que se conoce como *distribución de Rayleigh.*

Ejemplo 5.11 Un fabricante de lavadoras garantiza sus productos contra cualquier defecto durante el primer año de uso normal. El fabricante ha estimado un costo por reparación de $75 durante el periodo de garantía. Con base en la experiencia, se sabe que el tiempo en que ocurre la primera falla es una variable aletoria de Weibull con parámetros de forma y escala iguales a 2 y 40, respectivamente. Si el fabricante espera vender 100 mil unidades y si, para una misma unidad, se descuenta el valor de las reparaciones, se determina el costo esperado de la garantía para el fabricante.

Sea X la variable aleatoria que representa el tiempo que transcurre hasta que se presenta la primera descompostura. Por hipótesis, la función de densidad de probabilidad de X es:

$$f(x; 2, 40) = \frac{2}{40^2} x \exp[-(x/40)^2], \qquad x > 0.$$

La probabilidad de que la primera descompostura ocurra durante el periodo de garantía es igual a la probabilidad de que X sea menor o igual a 12. Mediante el empleo de (5.63), esta probabilidad es:

$$P(X \leq 12) = 1 - \exp[-(12/40)^2] = 0.0861.$$

Por lo tanto, si se supone que la operación de las lavadoras es independiente entre sí, se pueden esperar (100 000)(0.861) = 8610 de fallas durante el tiempo de garantía con un costo total de $645 750.

5.7 La distribución exponencial negativa

Se ha notado con anterioridad que la distribución exponencial (negativa) es un caso especial de los modelos de Weibull y gama. Ya que es un caso especial de la distribución gama (Erlang), la variable aleatoria exponencial es el tiempo que transcurre hasta que se da el primer evento de Poisson. Es decir, la distribución exponencial puede modelar el lapso entre dos eventos consecutivos de Poisson que ocurren de manera independiente y a una frecuencia constante. Esta distribución se emplea con bastante frecuencia con objeto de modelar problemas del tipo tiempo-falla y como modelo para el intervalo en problemas de líneas de espera. Posteriormente se demostrará que la distribución exponencial no tiene "memoria". Es decir, la probabilidad de ocurrencia de eventos presentes o futuros no depende de los que hayan ocurrido en el pasado. De esta forma, la probabilidad de que una unidad falle en un lapso específico depende nada más de la duración de éste, no del tiempo en que la unidad ha estado en operación.

Definición 5.6 Si una variable aleatoria X tiene una distribución exponencial, su función de densidad de probabilidad está dada por:

$$f(x;\theta) = \begin{cases} \dfrac{1}{\theta}\exp(-x/\theta) & x > 0, \quad \theta > 0, \\ 0 & \text{para cualquier otro valor.} \end{cases} \tag{5.71}$$

La distribución exponencial se caracteriza por un parámetro θ, que representa el lapso promedio de tiempo entre dos eventos independientes de Poisson. En el contexto de la confiabilidad, θ recibe el nombre de tiempo promedio entre fallas, y $1/\theta$ es la frecuencia de falla. La función de distribución acumulativa se obtiene directamente de los modelos de Weibull o de Erlang y está determinada por

$$P(X \leq x) = F(x;\theta) = 1 - \exp(-x/\theta). \tag{5.72}$$

Las expresiones para los valores cuantiles, momentos y factores de forma para esta distribución, se obtienen de las correspondientes expresiones para la distribución de Weibull con $\alpha = 1$. Esto es:

$$\begin{aligned} x_q &= \theta\ln[1/(1-q)], \\ E(X) &= \theta, \\ Var(X) &= \theta^2, \\ \alpha_3(X) &= 2, \text{ y} \\ \alpha_4(X) &= 9. \end{aligned}$$

En problemas de confiabilidad, generalmente el interés recae en determinar el tiempo de vida promedio de un componente o de un sistema de éstos. El problema esencial consiste en identificar la distribución de probabilidad de la variable aleatoria que, de manera adecuada, proporciona un modelo para el tiempo de falla. En esta línea, una cantidad muy útil es la función de confiabilidad.

Definición 5.7 Sea T una variable aleatoria que representa el tiempo de vida de un sistema y sea $f(t)$ la función de densidad de probabilidad de T. La función de confiabilidad del sistema a tiempo t, $R(t)$, es la probabilidad de que el lapso de duración del sistema sea mayor que un tiempo t dado. De acuerdo con lo anterior,

$$R(t) = P(T > t) = 1 - F(t), \qquad t > 0. \tag{5.73}$$

Otra cantidad muy útil para seleccionar una función de densidad de probabilidad para el lapso de vida medio de una unidad (o sistema) es la frecuencia de falla o función de riesgo, que se define de la siguiente forma:

Definición 5.8 Sean $f(t)$ y $R(t)$ las funciones de densidad de probabilidad y de confiabilidad, respectivamente, de una unidad en un tiempo dado t. La *frecuencia de falla* $h(t)$ se define como la proporción de unidades que fallan en el intervalo

$(t, t + dt)$ con respecto a las que siguen funcionando a tiempo t, y está determinada por:

$$h(t) = f(t)/R(t). \qquad (5.74)$$

Si se conoce la frecuencia de falla, es posible determinar la función de densidad de probabilidad de la variable aleatoria. Dado que $R(t) = 1 - F(t)$, mediante diferenciación con respecto a t, se tiene que $R'(t) = -F'(t)$; pero $F'(t) = f(t)$. Como resultado se tiene que la frecuencia de falla puede expresarse como:

$$h(t) = -R'(t)/R(t). \qquad (5.75)$$

Suponiendo que el sistema comenzó a funcionar en $t = 0$, $R(0) = 1$. Integrando ambos miembros de (5.75) desde 0 hasta t, se tiene:

$$\begin{aligned}\int_0^t h(x)dx &= -\int_0^t [R'(x)/R(x)]dx \\ &= -\ln[R(t)] + \ln[R(0)] \\ &= -\ln[R(t)],\end{aligned}$$

donde x es una variable muda de integración. Dado que:

$$-\ln[R(t)] = \int_0^t h(x)dx,$$

se tiene:

$$R(t) = \exp\left[-\int_0^t h(x)dx\right].$$

Mediante el empleo de (5.74), la función de densidad de probabilidad es:

$$f(t) = h(t)\exp\left[-\int_0^t h(x)dx\right]. \qquad (5.76)$$

Existen muchos fenómenos físicos de naturaleza aleatoria que muestran frecuencias de falla que tienen un parecido a "la curva de la tina de baño", tal y como se ilustra en la figura 5.9. En el intervalo de tiempo, de 0 a t_1, la frecuencia de falla es apreciable pero disminuye en valor debido al "síndrome de mortalidad infantil", mismo que sugiere que las primeras fallas pueden tener su origen en defectos de fabricación. Durante el intervalo de t_1 a t_2, $h(t)$ es casi constante, pero comienza a aumentar de valor después de t_2 por fallas debidas al desgaste de los componentes. Se puede imaginar una frecuencia de falla constante si los componentes se prueban inicialmente para detectar fallas por desgaste y se reemplazan antes de t_2.

Si la frecuencia de falla $1/\theta$, es constante, la función de densidad de probabilidad del tiempo de vida medio es la exponencial negativa. Esto es, si $h(t) = 1/\theta$, entonces de (5.76) se tiene:

$$\begin{aligned}f(t) &= \frac{1}{\theta}\exp\left[-\int_0^t \frac{1}{\theta}dx\right] \\ &= \frac{1}{\theta}\exp(-t/\theta).\end{aligned}$$

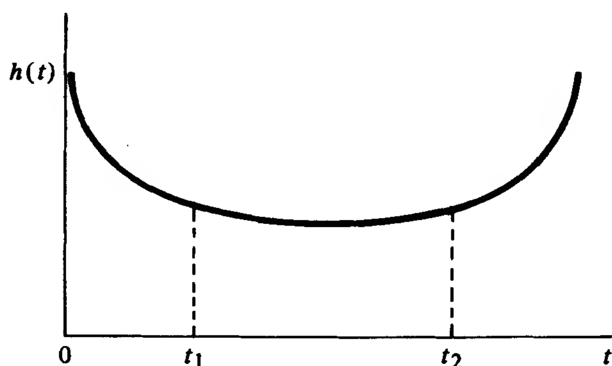


FIGURA 5.9 Función de frecuencia de falla típica

Nótese que la proposición inversa también es cierta; si el tiempo de vida medio se encuentra distribuido de manera exponencial, la frecuencia de falla es constante. Dado que la función de confiabilidad a tiempo t para un tiempo de vida medio distribuido exponencialmente es:

$$R(t) = \exp(-t/\theta), \qquad t > 0, \tag{5.77}$$

la frecuencia de falla está dada por:

$$\begin{aligned} h(t) &= \frac{(1/\theta)\exp(-t/\theta)}{\exp(-t/\theta)} \\ &= 1/\theta, \qquad t > 0. \end{aligned}$$

Una frecuencia de falla constante implica que la probabilidad de falla en un intervalo de tiempo determinado, depende de la duración de éste y no del tiempo en que el sistema ha estado operando. Esta última es la propiedad de "no memoria". A pesar de que el lapso de vida media no se encuentra distribuido de manera exponencial a lo largo de todo el periodo de funcionamiento del componente, el tiempo de operación de un sistema que contiene a éstos puede modelarse de manera adecuada por una distribución exponencial si se añade una selección inicial y una política de mantenimiento adecuada para los componentes.

Muchos investigadores proporcionan justificación empírica para la distribución exponencial en problemas de confiabilidad. El trabajo de Davis [3], quien demostró que el lapso de duración de ciertos componentes eléctricos puede modelarse de manera adecuada por una distribución exponencial, es típico en este sentido. Como ejemplo de este trabajo, la tabla 5.14 contiene una comparación entre las frecuencias observada y teórica para el tiempo de duración del bulbo V805. El tiempo de vida promedio para este bulbo, con base en los datos que se observaron fue de 179 horas. Al sustituir este valor de θ en (5.72), se pueden obtener las probabilidades teóricas para la distribución exponencial.

TABLA 5.14 Frecuencias observada y esperada para el bulbo V805

Tiempo de duración (horas)	*Frecuencia observada*	*Frecuencia relativa*	*Probabilidad del intervalo*	*Frecuencia esperada*
0– 80	317	0.3511	0.3604	325.4
80–160	230	0.2547	0.2305	208.2
160–240	118	0.1307	0.1474	133.1
240–320	93	0.1030	0.0943	85.2
320–400	49	0.0543	0.0603	54.5
400–480	33	0.0365	0.0386	34.8
480–560	17	0.0188	0.0247	22.3
560–700	26	0.0288	0.0238	21.5
700 o más	20	0.0221	0.0200	18.1
Totales	903	1.0000	1.0000	903.1

El argumento para emplear la distribución exponencial como modelo para el tiempo aleatorio en problemas de líneas de espera es similar al que se emplea en los lapsos de duración de un componente. Esto es, si un taller de reparación opera por un tiempo suficientemente largo para obtener una condición cercana al equilibrio, la probabilidad de hacer una reparación o que ésta se complete en un tiempo determinado, dependerá de este último, y no del que haya transcurrido en llevar a cabo la última reparación o el completarla.

A pesar de que la distribución exponencial negativa se emplea muchas veces para modelar la duración aleatoria de cierto componente, no es la distribución más apropiada, en el tiempo en que ocurrirá una falla, para todos los dispositivos. Existe una razón para creer que el lapso de tiempo que el componente tiene en operación afecta su duración. Los modelos más apropiados en estos casos son la distribución de Weibull o la de Erlang. Éstas exhiben frecuencias de falla crecientes, decrecientes o constantes dependiendo de cuándo los valores de los parámetros de forma son más grandes que, menores que, o iguales a uno, respectivamente. Por ejemplo, la función de confiabilidad para la distribución de Weibull está determinada por:

$$R(t) = \exp[-(t/\theta)^{\alpha}] \tag{5.78}$$

y la frecuencia de falla es:

$$h(t) = \alpha t^{\alpha-1}/\theta^{\alpha}. \tag{5.79}$$

Un ejemplo de sistema con una frecuencia de falla decreciente es aquél que mejora su funcionamiento con el paso del tiempo. Un ejemplo de este fenómeno es la duración de una empresa. Entre más tiempo tenga ésta operando con menor frecuencia se observará una falla en un intervalo dado de tiempo.

5.8 La distribución de una función de variable aleatoria

Uno de los ingredientes clave en inferencia estadística es la distribución de probabilidad de la "estadística" con base en la cual se formula la inferencia. Puesto que las

estadísticas son funciones de variables aleatorias, en muchas ocasiones es posible obtener sus distribuciones si se conocen las variables aleatorias sobre las que éstas se basan.

En esta sección se examinará una técnica para determinar la distribución de una función de variable aleatoria, considerando el caso de una variable aleatoria continua. Sea X una variable aleatoria con función de densidad de probabilidad $f_X(x)$, y sea $Y = g(X)$ una función definida de X. Supóngase que es posible resolver $y = g(x)$ para x obteniendo de esta forma la función inversa $x = g^{-1}(y)$. Si $g(x)$ y $g^{-1}(y)$ son funciones univaluadas de x y y, respectivamente, se dice que la transformación es uno a uno. Esto es, a cada punto en el espacio muestral de X le corresponde un punto único del espacio muestral de Y y viceversa. Si se supone la existencia de una transformación uno a uno y además que $y = g(x)$ es una función creciente y diferenciable de x, se puede determinar la función de densidad de probabilidad de X en la siguiente forma:

Debido a la existencia de una transformación uno a uno:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P[g(X) \leq y] \\ &= P[X \leq g^{-1}(y)]. \end{aligned}$$

Entonces:

$$F_Y(y) = F_X[g^{-1}(y)]. \tag{5.80}$$

Al establecer la diferencia (5.80) con respecto a y y mediante el empleo de la regla de la cadena, se tiene:

$$\begin{aligned} f_Y(y) &= \frac{dF_X[g^{-1}(y)]}{dx} \cdot \frac{dx}{dy} \\ &= f_X[g^{-1}(y)]\frac{dx}{dy}. \end{aligned} \tag{5.81}$$

Si $g(x)$ es una función decreciente de x, el resultado que se obtiene es el mismo con excepción de que la derivada de una función decreciente es negativa. De esta manera se puede formular la siguiente proposición:

Teorema 5.2 Sea X una variable aleatoria continua con función de densidad de probabilidad $f_X(x)$ y defínase $Y = g(X)$. Si $y = g(x)$ y $x = g^{-1}(y)$ son funciones univaluadas, continuas y diferenciables y si $y = g(x)$ es una función creciente o decreciente de x, la función de densidad de probabilidad de Y está determinada por:

$$f_Y(y) = f_X[g^{-1}(y)]\left|\frac{dx}{dy}\right|, \tag{5.82}$$

en donde la cantidad $J = |dx/dy|$ recibe el nombre de *Jacobiano* de la transformación.

El teorema 5.2 se obtiene a partir de una técnica de cambio de variable en una integral definida, que ya se empleó en varias ocasiones.

Sea X una variable aleatoria continua con una función de densidad de probabilidad $f(x;\ \mu,\ \theta,\ \alpha)$, donde μ, θ, y α son los parámetros de localización, escala y forma respectivamente. El efecto del parámetro de forma α puede hacerse más claro si se considera la distribución de la variable aleatoria estandarizada $Y = (X - \mu)/\theta$, la cual no contiene a μ y θ. Mediante el empleo de (5.82), la función de densidad de probabilidad de Y es:

$$f_Y(y) = \theta f_X(\theta y + \mu), \tag{5.83}$$

ya que la relación inversa es $x = \theta y + \mu$ y el Jacobiano está dado por $dx/dy = \theta$. En particular, sea X una variable aleatoria con distribución gama y cuya función de densidad se establece por (5.45). La función de densidad de $Y = X/\theta$ es:

$$f_G(y;\alpha) = \frac{1}{\Gamma(\alpha)}\, y^{\alpha-1}\exp(-y), \qquad y > 0. \tag{5.84}$$

De manera similar, si X es una variable aleatoria de Weibull con función de densidad de probabilidad dada por (5.61), la densidad de $Y = X/\theta$ es:

$$f_W(y;\alpha) = \alpha y^{\alpha-1}\exp(-y^{\alpha}), \qquad y > 0. \tag{5.85}$$

Si no existe un parámetro de forma y si μ y θ son la media y la desviación estándar de X, respectivamente, entonces (5.83) dará origen a una función de densidad libre de parámetros con media cero y desviación estándar uno. Un ejemplo de lo anterior es la función de densidad de probabilidad normal estandarizada.

Ejemplo 5.12. Si la variable aleatoria X se encuentra distribuida de manera uniforme en el intervalo $(0, \pi)$, debe obtenerse la función de densidad de probabilidad de la función $Y = c \operatorname{sen}(X)$, para cualquier constante positiva c.

Nótese que la relación $y = c \operatorname{sen}(x)$ es una función estrictamente creciente de x en el intervalo $(0,\ \pi/2)$ y estrictamente decreciente en el intervalo $(\pi/2,\ \pi)$. Cuando la relación funcional es creciente en alguna parte del dominio de la variable aleatoria original y decreciente para el resto, la función de densidad de probabilidad de interés puede obtenerse al tratar cada parte de manera separada y sumar los resultados. De acuerdo con lo anterior, los intervalos $(0,\ \pi/2)$ y $(\pi/2,\ \pi)$ deben manejarse en forma separada.

La relación inversa es:

$$x = \operatorname{sen}^{-1}(y/c),$$

y el Jacobiano de la transformación es:

$$\left|\frac{dx}{dy}\right| = \frac{1}{c}\left[1 - \left(\frac{y}{c}\right)^2\right]^{-1/2}$$
$$= (c^2 - y^2)^{-1/2}.$$

Dado que la densidad de X es:

$$f(x) = 1/\pi \qquad 0 \leq x \leq \pi,$$

para el intervalo $(0, \pi/2)$,

$$f_1(y) = \frac{1}{\pi}(c^2 - y^2)^{-1/2} \qquad 0 \leq y \leq c,$$

y para el intervalo $(\pi/2, \pi)$,

$$f_2(y) = \frac{1}{\pi}(c^2 - y^2)^{-1/2} \qquad 0 \leq y \leq c.$$

La función de densidad de probabilidad de Y es:

$$f_Y(y) = f_1(y) + f_2(y)$$

$$= \frac{2}{\pi}(c^2 - y^2)^{-1/2}, \qquad 0 \leq y \leq c. \tag{5.86}$$

Ejemplo 5.13 Sea X una variable aleatoria distribuida normalmente con media μ y desviación estándar σ. Obtener la función de densidad de probabilidad de $Y = \exp(X)$.

La relación $y = \exp(x)$ es una función creciente y diferenciable de x. La relación inversa es $x = \ln(y)$, y el Jacobiano es $dx/dy = 1/y$. Por lo tanto, la densidad de Y es:

$$f_Y(y; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma y} \exp\left\{ -\frac{1}{2}\left[\frac{\ln(y) - \mu}{\sigma}\right]^2 \right\}, \qquad y > 0. \tag{5.87}$$

La expresión dada en (5.87) es la función de densidad de probabilidad de lo que se conoce como el *modelo log-normal*. A pesar de que los parámetros de la densidad log-normal son las cantidades μ y σ, éstas no representan parámetros de localización o escala. Más bien son la media y la desviación estándar de la correspondiente variable aleatoria normal. Mientras que la variable aleatoria normal se considera, en muchas ocasiones, como la representante del efecto aditivo de muchos errores físicos pequeños, la variable aleatoria log-normal representa el efecto multiplicativo de éstos. La distribución log-normal se emplea en una gran variedad de aplicaciones que incluyen el problema de evaluar los efectos de la fatiga sobre materiales. Véase [1] para una presentación detallada de esta distribución.

Existe otro método para determinar la distribución de una función de variable aleatoria que emplea la función generadora de momentos. Recuérdese que esta función, si existe, determina de manera unívoca una distribución de probabilidad. De esta manera, si se encuentra que una variable aleatoria tiene la misma función generadora de momentos que la de una distribución conocida, entonces la función de variable aleatoria tiene la misma distribución.

Ejemplo 5.14 Sea Z una variable aleatoria distribuida normalmente con media cero y desviación estándar uno. Demostrar que la distribución de:

$$Y = Z^2$$

es una distribución chi-cuadrado con un grado de libertad.

Por definición, la función que genera momentos de Z^2 es:

$$\begin{aligned} m_{Z^2}(t) = E[\exp(tZ^2)] &= \int_{-\infty}^{\infty} \exp(tz^2)f(z)dz \\ &= (2\pi)^{-1/2}\int_{-\infty}^{\infty} \exp(tz^2)\exp(-z^2/2)dz \\ &= (2\pi)^{-1/2}\int_{-\infty}^{\infty} \exp[-(z^2/2)(1-2t)]dz \\ &= (2\pi)^{-1/2}\int_{-\infty}^{\infty} \exp\left[-\frac{z^2}{2(1-2t)^{-1}}\right]dz. \end{aligned}$$

Nótese que, excepto por una constante, el integrando de la última integral es igual al de la función de densidad de probabilidad de una variable aleatoria normal con media cero y varianza $(1 - 2t)^{-1}$. Para hacer el integrando igual a una distribución normal con media cero y varianza $(1 - 2t)^{-1}$, se multiplica tanto el numerador como el denominador por la desviación estándar $(1 - 2t)^{-1/2}$, que no es otra cosa más que multiplicar la expresión por uno. De esta forma,

$$\begin{aligned} m_{Z^2}(t) &= \frac{1}{(1-2t)^{1/2}}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,(1-2t)^{-1/2}}\exp\left[-\frac{z^2}{2(1-2t)^{-1}}\right]dz \\ &= (1-2t)^{-1/2}, \end{aligned}$$

dado que el integrando es una función de densidad de probabilidad normal y por definición, la integral desde $-\infty$ a ∞, es uno. La función generadora de momentos de $Y = Z^2$ es idéntica a la de la distribución chi-cuadrado con $\nu = 1$ grados de libertad; (véase (5.60)). Por lo tanto, el cuadrado de la variable aleatoria normal estándar tiene una distribución chi-cuadrado con un grado de libertad.

5.9 Conceptos básicos en la generación de números aleatorios por computadora

Desde el advenimiento de los sistemas de cómputo de gran escala, los experimentos de simulación se han convertido en técnicas muy útiles para el análisis de sistemas complejos que, muchas veces, se constituyen por muchos componentes interdependientes. En la simulación de estos sistemas surge la necesidad de simular fenómenos aleatorios que son característicos de un sistema en especial. Por ejemplo, si un banco desea examinar su sistema de servicios al cliente, debe simular el flujo de clientes al

banco, así como también el tiempo necesario para llevar a efecto cada operación bancaria y estos sucesos constituyen eventos aleatorios.

Para atacar este tipo de problemas se supone, en general, una distribución de probabilidad apropiada para cada fenómeno y se genera una secuencia de valores para la correspondiente variable aleatoria por computadora. Puesto que estas secuencias se generan mediante el empleo de algoritmos numéricos que pueden repetirse exactamente, estas secuencias de números no constituyen, en un sentido estricto, números aleatorios. Sin embargo, estas secuencias exhiben suficientes propiedades aleatorias para emplearse con éxito en muchas aplicaciones.

El propósito de esta sección no es estudiar las propiedades de los números aleatorios generados por computadora ni determinar la forma más eficiente de hacerlo. Más bien el propósito es familiarizar al lector con las posibles formas de generar números aleatorios a partir de alguna de las distribuciones de probabilidad, discretas y continuas, que se han estudiado.

La distribución uniforme sobre el intervalo (0, 1) juega un papel muy importante en la generación de números aleatorios por computadora. Para finalizar se establece y demuestra el siguiente teorema:

Teorema 5.3 Para cualquier variable aleatoria continua X, la función de distribución acumulativa $F(x; \theta)$ con parámetro θ se puede representar por una variable aleatoria U, la cual se encuentra uniformemente distribuida sobre el intervalo unitario.

Demostración: Dado que por definición la función de distribución acumulativa de X está dada por:

$$F(x; \theta) = \int_{-\infty}^{x} f(t; \theta)dt,$$

a cada valor de x le corresponde un valor de $F(x; \theta)$ que necesariamente se encuentra en el intervalo (0, 1). Además, $F(X; \theta)$ también es una variable aleatoria en virtud de la aleatoriedad de X. Para cada valor u de la variable aleatoria U, la función $u = F(x; \theta)$ define una correspondencia uno a uno entre U y X siendo la relación inversa $x = F^{-1}(u)$. Al tener $du = dF(x; \theta) = f(x; \theta)dx$, el Jacobiano de la transformación es:

$$J = \left|\frac{dx}{du}\right| = [f(x; \theta)]^{-1} = [f(F^{-1}(u); \theta)]^{-1}.$$

La función de densidad de probabilidad de la variable aleatoria U, mediante el empleo de (5.82), es:

$$\begin{aligned} g(u) &= f(F^{-1}(u); \theta)[f(F^{-1}(u); \theta)]^{-1} \\ &= 1, \quad 0 \leq u \leq 1. \end{aligned}$$

La esencia del teorema 5.3 recae en el hecho de que, para muchos casos, es posible determinar de manera directa el valor de x que corresponde al valor de u de las variables aleatoria X y U, respectivamente, de manera tal que $F(x; \theta) = u$. Por esta ra-

zón todos los sistemas de cómputo tienen en su estructura la capacidad de generar valores aleatorios a partir de una distribución uniforme sobre el intervalo unitario (0, 1). De hecho, muchos paquetes estadísticos para computadora, como SAS, SPSS y IMSL, proporcionan al usuario la oportunidad de generar números aleatorios a partir de una distribución dada. Se ilustrará el uso del teorema 5.3 en la generación de números aleatorios para algunas distribuciones de probabilidad específicas.

5.9.1 Distribución uniforme sobre el intervalo (a, b)

La función de densidad de probabilidad es:

$$f(x; a, b) = 1/(b - a), \qquad a \le x \le b.$$

Para generar un número aleatorio x, $a \le x \le b$, primero se genera un valor aleatorio u a partir de (0, 1), se iguala a la función de distribución acumulativa, se integra y se resuelve para el límite superior x. De esta forma:

$$(b - a)^{-1} \int_a^x dt = u$$

$$\frac{x - a}{b - a} = u,$$

o

$$x = u(b - a) + a, \qquad a \le x \le b \tag{5.88}$$

5.9.2 La distribución de Weibull

La función de densidad de probabilidad es:

$$f(x; \alpha, \theta) = \frac{\alpha}{\theta^\alpha} x^{\alpha-1} \exp[-(x/\theta)^\alpha], \qquad x > 0.$$

Para generar números aleatorios de Weibull $x > 0$, se resuelve la ecuación

$$\frac{\alpha}{\theta^\alpha} \int_0^x t^{\alpha-1} \exp[-(t/\theta)^\alpha] dt = u$$

$$\left(\frac{\alpha}{\theta^\alpha}\right)\left(-\frac{\theta^\alpha}{\alpha}\right) \exp[-(t/\theta)^\alpha]\Big|_0^x = u$$

o

$$1 - \exp[-(x/\theta)^\alpha] = u,$$

y

$$x = \theta\left[\ln\left(\frac{1}{1 - u}\right)\right]^{1/\alpha}. \tag{5.89}$$

Dado que para $\alpha = 1$, la distribución de Weibull se reduce a la exponencial, pueden generarse números aleatorios para una distribución exponencial mediante (5.89) con $\alpha = 1$.

5.9.3 La distribución de Erlang

La función de densidad de probabilidad es:

$$f(x; \alpha, \theta) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1}\exp(-x/\theta), \qquad x > 0,$$

en donde α es un entero positivo. Recuérdese que la variable aleatoria de Erlang es la suma de α variables aleatorias independientes distribuidas exponencialmente. Por lo tanto, un número aleatorio de Erlang es la suma de α valores aleatorios exponenciales, en donde cada valor se genera mediante (5.89).

5.9.4 La distribución normal

La función de distribución acumulativa normal es:

$$\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x} \exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right]dt = u$$

no puede resolverse, en forma cerrada, para x. De manera alternativa, puede demostrarse que si U_1 y U_2 son dos variables aleatorias independientes con distribución uniforme sobre el intervalo unitario, entonces

$$\begin{aligned} Z_1 &= (-2\ \ln U_1)^{1/2}\operatorname{sen}(2\pi U_2) \quad \text{y} \\ Z_2 &= (-2\ \ln U_1)^{1/2}\cos(2\pi U_2) \end{aligned} \tag{5.90}$$

son dos variables aleatorias normales estandarizadas e independientes.

5.9.5 La distribución binomial

Para generar números aleatorios a partir de una distribución binomial con función de probabilidad se considerará lo siguiente: la variable aleatoria binomial es vista como la suma de n resultados de un proceso de Bernoulli descrito por:

$$p(x; n, p) = \frac{n!}{(n-x)!\,x!} p^x(1-p)^{n-x}, \qquad x = 0, 1, 2, \ldots, n$$

$$Y = \begin{cases} 1 & \text{con probabilidad } p \\ 0 & \text{con probabilidad } (1-p). \end{cases}$$

Se puede obtener un número aleatorio binomial mediante la suma de n de los valores de la variable aleatoria Y, en donde cada valor se determina mediante:

$$y = \begin{cases} 1 & \text{si } 0 \le u \le p \\ 0 & \text{si } p < u \le 1, \end{cases} \tag{5.91}$$

donde u es un número aleatorio uniforme sobre el intervalo unitario. Esto es, se generan n números aleatorios a partir del intervalo unitario, se convierten a unos y ceros de acuerdo con (5.91) y la suma de los unos en esta secuencia es el número aleatorio binomial.

5.9.6 La distribución de Poisson

Recuérdese que la probabilidad de tener x ocurrencias en un intervalo de tiempo t está definida por:

$$p(x; t) = \frac{(\nu t)^x \exp(-\nu t)}{x!}, \qquad x = 0, 1, 2, \ldots,$$

donde ν es la frecuencia constante de ocurrencia, y $\lambda = \nu t$ es el número promedio de éstas. Como la ocurrencia en el tiempo de dos eventos independientes de Poisson se encuentra distribuida exponencialmente, se puede generar un número aleatorio de Poisson x mediante la generación sucesiva de números aleatorios exponenciales por (5.89) para $\alpha = 1$. El proceso se continúa hasta que la suma de los valores $x + 1$ sea mayor que el intervalo de tiempo t. Por lo tanto, el número aleatorio de Poisson es x.

Referencias

1. A. Aitchison and J. A. C. Brown, *The log-normal distribution*, Cambridge Univ. Press, Cambridge, England, 1957.
2. K. V. Bury, *Statistical models in applied science*, Wiley, New York, 1975.
3. D. J. Davis, *An analysis of some failure data*, J. of the Amer. Statistical Assoc. **47** (1952), 113–150.
4. D. G. Malcolm, J. H. Roseboom, E. C. Clark, and W. Fazar, *Application of a technique for research and development program evaluation*, Operations Research **7** (1959), 646–658.
5. E. S. Pearson and H. O. Hartley, Eds., *Biometrika tables for statisticians*, Vol. I, 3rd ed., Cambridge Univ. Press, Cambridge, England, 1966.
6. K. Pearson, *Tables of the incomplete beta function*, Biometrika Office, University College, London, 1948.
7. K. Pearson, *Tables of the incomplete gamma function*, Biometrika Office, University College, London, 1957.
8. H. C. S. Thom, *Direct and inverse tables of the gamma distribution*, Environmental Data Service, Silver Spring, Md., 1968.
9. W. Weibull, *A statistical distribution function of wide applicability*, J. Appl. Mech. **18** (1951), 293–302.

Ejercicios

5.1. En la misma gráfica, dibujar las distribuciones normales $N(0, 5)$ y $N(0, 4)$

5.2. Sea $X \sim N(50, 10)$. Determinar las siguientes probabilidades:

a) $P(X < 40)$ *d)* $P(X > 35)$
b) $P(X < 65)$ *e)* $P(40 < X < 45)$
c) $P(X > 55)$ *f)* $P(38 < X < 62)$

5.3. Sea $X \sim N(200, 20)$. Determinar las siguientes probabilidades:

a) $P(185 < X < 210)$ *c)* $P(X > 240)$
b) $P(215 < X < 250)$ *d)* $P(X > 178)$

5.4. Sea $X \sim N(-25, 10)$. Encontrar los valores de x que corresponden a las siguientes probabilidades:

a) $P(X < x) = 0.1251$ *c)* $P(X > x) = 0.3859$
b) $P(X < x) = 0.9382$ *d)* $P(X > x) = 0.8340$

5.5. Sea $X \sim N(10, 5)$. Encontrar los valores de x que corresponden a las siguientes probabilidades:

a) $P(X < x) = 0.05$ *d)* $P(X < x) = 0.01$
b) $P(X < x) = 0.95$ *e)* $P(X < x) = 0.025$
c) $P(X < x) = 0.99$ *f)* $P(X < x) = 0.975$

5.6. Sea $X \sim N(\mu, \sigma)$. Determinar la media y la varianza de X si los cuantiles son $x_{0.4} = 50$ y $x_{0.8} = 100$

5.7. Una universidad espera recibir, para el siguiente año escolar, 16 000 solicitudes de ingreso al primer año de licenciatura. Se supone que las calificaciones obtenidas por los aspirantes en la prueba SAT se pueden calcular, de manera adecuada, por una distribución normal con media 950 y desviación estándar 100. Si la universidad decide admitir al 25% de todos los aspirantes que obtengan las calificaciones más altas en la prueba SAT, ¿cuál es la mínima calificación que es necesario obtener en esta prueba, para ser admitido por la universidad?

5.8. Una fábrica produce pistones cuyos diámetros se encuentran adecuadamente clasificados por una distribución normal con un diámetro promedio de 5 cm y una desviación estándar igual a 0.001 cm. Para que un pistón sirva, su diámetro debe encontrarse entre 4.998 y 5.002 cm. Si el diámetro del pistón es menor que 4.998 se desecha; si es mayor que 5.002 el pistón puede reprocesarse. ¿Qué porcentaje de pistones servirá? ¿Qué porcentaje será desechado? ¿Qué porcentaje será reprocesado?

5.9. La demanda mensual de cierto producto A tiene una distribución normal con una media de 200 unidades y desviación estándar igual a 40 unidades. La demanda de otro producto B también tiene una distribución normal con media de 500 unidades y desviación estándar igual a 80 unidades. Un comerciante que vende estos productos tiene en su almacén 280 unidades de A y 650 de B al comienzo de un mes, ¿cuál es la probabilidad de que, en el mes, se vendan todas las unidades de ambos productos? Puede suponerse independencia entre ambos eventos.

5.10. El peso de cereal que contiene una caja se aproxima a una distribución normal con una media de 600 gramos. El proceso de llenado de las cajas está diseñado para que de entre 100 cajas, el peso de una se encuentre fuera del intervalo 590-610 gramos. ¿Cuál es el valor máximo de la desviación estándar para alcanzar este requerimiento?

5.11. En una tienda de descuento la demanda diaria de acumuladores para automóvil se calcula mediante una distribución normal con una media de 50 acumuladores que tienen

una desviación estándar de 10. En dos días consecutivos se venden 80 y 75 acumuladores respectivamente. Si estos días son típicos, ¿qué tan probable es, bajo las suposiciones dadas, vender 80 o más y 75 o más acumuladores?

5.12. Un fabricante de aviones desea obtener remaches para montar los propulsores de sus aviones. El esfuerzo a la tensión mínimo necesario de cada remache es de 25 000 lb. Se pide a tres fabricantes de remaches (A, B y C) que proporcionen toda la información pertinente con respecto a los remaches que producen. Los tres fabricantes aseguran que la resistencia a la tensión de sus remaches se encuentra distribuida, de manera aproximada, normalmente con un valor medio de 28 000, 30 000 y 29 000 lb, respectivamente.

a) ¿Tiene el fabricante la suficiente información para hacer una selección? ¿Por qué?
b) Supóngase que las desviaciones estándar para A, B y C son 1 000, 1800 y 1200, respectivamente. ¿Cuál es la probabilidad de que un remache producido ya sea por A, B o C no reúna los requisitos mínimos?
c) Si usted fuera el fabricante de aviones, ¿podría elegir entre A, B y C, con base en su respuesta al inciso *b)*? ¿Por qué?

5.13. Un fabricante de escapes para automóviles desea garantizar su producto durante un periodo igual al de la duración del vehículo. El fabricante supone que el tiempo de duración de su producto es una variable aleatoria con una distribución normal, con una vida promedio de tres años y una desviación estándar de seis meses. Si el costo de reemplazo por unidad es de $10, ¿cuál puede ser el costo total de reemplazo para los primeros dos años, si se instalan 1 000 000 unidades?

5.14. El tiempo necesario para armar cierta unidad es una variable aleatoria normalmente distribuida con una media de 30 minutos y desviación estándar igual a dos minutos. Determinar el tiempo de armado de manera tal que la probabilidad de exceder éste sea de 0.02.

5.15. Un periódico llevó a cabo una encuesta entre 400 personas seleccionadas aleatoriamente, en un estado, sobre el control de armas. De las 400 personas, 220 se pronunciaron en favor de un estricto control.

a) ¿Qué tan probable resulta el hecho de tener 220 o más personas a favor del control de armas, si la población en este estado se encuentra dividida en opinión de igual manera?
b) Supóngase que se encuesta a 2000 personas teniendo la misma proporción de éstas a favor del control de armas, que la del inciso anterior. ¿Cómo cambiaría su respuesta al inciso *a)*?
c) Si el número de personas encuestadas es de 10 000, ¿cuál es la probabilidad de tener una ocurrencia diferente a la del inciso *b*)?

5.16. Una prueba de opción múltiple contiene 25 preguntas y cada una de éstas cinco opciones. ¿Cuál es la probabilidad de que, al contestar de manera aleatoria cada pregunta, más de la mitad de las respuestas sea incorrecta?

5.17. Una organización llevó a cabo una encuesta entre 1 600 personas, seleccionadas de manera aleatoria de toda la población del país, para conocer su opinión con respecto a la seguridad en las plantas de energía nuclear. De este grupo, el 60% opinó que las plantas de energía nuclear tienen muy poca seguridad. Con base en estos resultados ¿existe alguna razón para dudar que la población en general tiene una opinión neutral con respecto a este asunto?

5.18. Sea X una variable aleatoria distribuida binomialmente.

a) Para $n = 15, p = 0.25$ y $n = 15$ y $p = 0.5$, calcular las siguientes probabilidades: $P(X = 8)$, $P(X \leq 3)$, $P(X \leq 7)$, $P(X \geq 9)$, y $P(X \geq 12)$.
b) Aproxímense los valores de las probabilidades anteriores mediante el empleo de la distribución normal.
c) Repetir los incisos *a)* y *b)* para $n = 25$ y comparar los resultados.

5.19. Sea X una variable aleatoria con distribución uniforme sobre el intervalo (a, b).

a) ¿Cuál es la probabilidad de que X tome un valor que se encuentre a una desviación estándar de la media?
b) ¿Puede tomar X un valor que se encuentre a dos desviaciones estándar de la media?

5.20. Sea X una variable aleatoria con distribución uniforme sobre el intervalo (a, b). ¿Cuál es la máxima distancia, en términos de la desviación estándar, a la que puede encontrarse un valor X a partir de la media?

5.21. Sea X una variable aleatoria con distribución uniforme sobre el intervalo (a, b). Si $E(X) = 10$ y $Var(X) = 12$, encontrar los valores de a y de b.

5.22. Supóngase que la concentración de cierto contaminante se encuentra distribuida de manera uniforme en el intervalo de 4 a 20 ppm (partes por millón). Si se considera como tóxica una concentración de 15 ppm o más, ¿cuál es la probabilidad de que al tomarse una muestra la concentración de ésta sea tóxica?

5.23. Sea X una variable aleatoria con distribución beta y parámetros $\alpha = 3$ y $\beta = 1$.

a) Graficar la función de densidad de probabilidad.
b) Obtener la media, la varianza, la desviación media, el coeficiente de asimetría y la curtosis relativa.
c) ¿Cuál es la probabilidad de que X tome un valor que se encuentre dentro de una desviación estándar a partir de la media? ¿A dos desviaciones estándar?
d) Determinar los cuantiles de esta distribución.

5.24. Si los parámetros de la distribución beta son enteros, puede demostrarse que la función de distribución acumulativa beta se encuentra relacionada con la distribución binomial en la siguiente forma:

$$P(X < p) = I_p(\alpha, \beta) = \sum_{y=\alpha}^{n} \frac{n!}{(n-y)!y!} p^y (1-p)^{n-y},$$

en donde $n = \alpha + \beta - 1$ y $0 < p < 1$. Si X es una variable aleatoria con una distribución beta con parámetros $\alpha = 2$ y $\beta = 3$, emplear la relación anterior para obtener $P(X < 0.1)$, $P(X < 0.25)$, y $P(X < 0.5)$.

5.25. Tomando como referencia el ejercicio anterior, determinar la probabilidad de que X tome un valor que se encuentre dentro de un intervalo igual a una desviación estándar de la media y, posteriormente, de un intervalo igual a dos desviaciones estándar.

5.26. La proporción de unidades defectuosas en un proceso de fabricación es una variable aleatoria que se encuentra aproximada por una distribución beta con $\alpha = 1$ y $\beta = 20$.

a) ¿Cuál es el valor de la media y de la desviación estándar?
b) ¿Cuál es la probabilidad de que la proporción de artículos defectuosos sea mayor que un 10%? ¿Mayor que un 15%?

5.27. Aproxime su respuesta al inciso *b)* del ejercicio anterior mediante el empleo de la aproximación normal dada por la expresión (5.44).

5.28. La competencia en el mercado de una compañía de computadoras varía de manera aleatoria de acuerdo con una distribución beta con $\alpha = 10$ y $\beta = 6$.

a) Graficar la función de densidad de probabilidad.
b) Encontrar la media y la desviación estándar.
c) Obtener la probabilidad de que la competencia en el mercado sea menor que la media.
d) Encontrar la probabilidad de que la competencia en el mercado se encuentre dentro de una desviación estándar de la media y, posteriormente, de un intervalo igual a dos desviaciones estándar de la media.

5.29. Sea X una variable aleatoria con distribución gama con $\alpha = 2$ y $\theta = 50$.

a) ¿Cuál es la probabilidad de que X tome un valor menor al valor de la media?
b) ¿Cuál es la probabilidad de que X tome un valor mayor de dos desviaciones estándar con respecto a la media?
c) ¿Cuál es la probabilidad de que X tome un valor menor al de su moda?

5.30. Sea X una variable aleatoria con distribución gama y $\alpha = 2$ y $\theta = 100$.
a) Graficar la función de densidad de probabilidad.
b) Encontrar la probabilidad de que, primero, X tome un valor dentro de un intervalo igual a una desviación estándar de la media y, posteriormente, de un intervalo igual a dos desviaciones estándar de la media.
c) ¿Cómo cambiarían sus respuestas a la parte *b)* si $\theta = 200$?

5.31. La edad a la que un hombre contrae matrimonio por primera vez es una variable aleatoria con distribución gama. Si la edad promedio es de 30 años y lo más común es que el hombre se case a los 22 años, encontrar los valores de los parámetros α y θ, para esta distribución.

5.32. La información que a continuación se presenta es una tabulación parcial de la función gama incompleta tal como se encuentra definida por (5.55) para $\alpha = 16$.

u	2	2.5	3.0	3.5	4.0	4.5
$I(u, 15)$	0.0082	0.0487	0.1556	0.3306	0.5333	0.7133

u	5.0	5.5	6.0	6.5	7.0
$I(u, 15)$	0.8435	0.9231	0.9656	0.9858	0.9946

Para $\theta = 10$, comparar estas probabilidades con las que se proporcionaron al emplear una aproximación normal.

5.33. Mediante el empleo de la función generadora de momentos de la distribución gama, encontrar expresiones para la media y la varianza.

5.34. La duración de cierto componente es una variable aleatoria con distribución gama y parámetro $\alpha = 2$.
a) Obtener la función de confiabilidad.
b) Para $\theta = 20$, obtener la frecuencia de falla y graficarla como una función de t.
c) Si $\theta = 20$, ¿cuál es la confiabilidad del componente en $t = 80$?

5.35. Para armar un artículo se necesitan cuatro etapas. Si el tiempo total necesario para armar un artículo, en horas, es una variable aleatoria con distribución gama y parámetro de escala $\theta = 2$, ¿cuál es la probabilidad de armar un artículo en menos de 15 horas?

5.36. Sea X una variable aleatoria con distribución de Weibull y parámetros $\alpha = 2$ y $\theta = 20$.

a) Graficar la función de densidad de probabilidad.
b) Obtener la probabilidad de que X tome un valor mayor que la media.
c) Obtener la probabilidad de que X tome un valor que se encuentre en un intervalo igual a una desviación estándar, y después en un intervalo igual a dos desviaciones estándar de la media.

5.37. El tiempo de duración de un sistema se encuentra aproximado por una distribución de Weibull con $\alpha = 2$ y $\theta = 50$.
a) Obtener la media y los deciles de esta distribución.
b) Obtener la confiabilidad de este sistema en $t = 75$.

5.38. Un sistema está formado por dos componentes independientes A y B. El sistema permanecerá operando mientras uno o ambos componentes funcionen. Si el tiempo de vida de la componente A es una variable aleatoria de Weibull con $\alpha = 1/2$ y $\theta = 10$, y si el tiempo de vida de B es también una variable de Weibull con $\alpha = 2$ y $\theta = 12$. ¿cuál es la probabilidad de que el sistema trabaje más de 20 horas?

5.39. Sea X una variable aleatoria con distribución exponencial.
a) ¿Cuál es la probabilidad de que X tome un valor mayor que la media?
b) Cuáles son las probabilidades de que X tome un valor que se encuentre en un intervalo igual a una desviación estándar, primero, y en un intervalo igual a dos desviaciones estándar de la media?

5.40. Si la frecuencia con que falla un componente es constante y la confiabilidad de éste tiene un valor en $t = 55$ de 0.4,

a) Obtener la función de densidad de probabilidad.
b) Obtener la confiabilidad del componente para $t = 100$.

5.41. Un dispositivo tiene una frecuencia de falla constante $h(t) = 10^{-2}$ por hora.

a) ¿Cuál es la confiabilidad del dispositivo para $t = 200$ horas?
b) Si 500 de estos dispositivos fallan de manera independiente, ¿cuál es el número esperado de fallas entre éstos, después de 200 horas?

5.42. El compresor de una unidad de aire acondicionado tiene una frecuencia de falla $h(t) = 2 \times 10^{-8}t$ por hora.

a) ¿Cuál es la función de confiabilidad del compresor?
b) ¿Cuál es la confiabilidad del compresor para $t = 15\,000$ horas?
c) ¿Cuál es la vida media del compresor?
d) ¿Cuál es la mediana de su duración?

5.43. Sea X una variable aleatoria con distribución uniforme en el intervalo (0, 1). Demostrar que la variable aleatoria $Y = -2 \ln(X)$ tiene una distribución chi-cuadrado con dos grados de libertad.

5.44. Si X es una variable aleatoria con una distribución exponencial y parámetro θ, obtener la distribución de $Y = (X - \theta)/\theta$.

5.45. Si X es una variable aleatoria con una distribución de Weibull y parámetros α y θ, obtener la distribución de $Y = X^{\alpha}$.

5.46. Seleccione una distribución de probabilidad discreta y una continua de la sección 5.9 y genere dos muestras aleatorias de 50 números aleatorios cada una. Para cada caso agru-

pe los datos y obtenga las frecuencias relativas. Calcule la media y la desviación estándar de cada una de las muestras y compare los resultados con los que se obtienen de manera teórica.

APÉNDICE

Demostración de que la expresión (5.1) es una función de densidad de probabilidad.

El que la función sea no negativa se satisface, ya que $f(x; \mu, \sigma) > 0$ para $-\infty < x < \infty$, $-\infty < \mu < \infty$ y $\sigma > 0$. Para demostrar que:

$$\int_{-\infty}^{\infty} f(x; \mu, \sigma)dx = 1,$$

sea:

$$I = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} \exp\left[-(x-\mu)^2/2\sigma^2\right]dx$$

el valor de la integral y aplíquese la transformación lineal $y = (x - \mu)/\sigma$ de manera tal que $x = \sigma y + \mu$ y $dx = \sigma dy$. Esto da como resultado:

$$I = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp(-y^2/2)dy.$$

Si puede demostrarse que $I^2 = 1$, puede deducirse que $I = 1$ puesto que $f(x; \mu, \sigma)$ tiene una valor positivo. De acuerdo con lo anterior:

$$I^2 = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp(-y^2/2)dy \cdot \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp(-z^2/2)dz$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left[-(y^2 + z^2)/2\right]dydz,$$

en donde se ha escrito el producto de las dos integrales como una doble integral ya que las funciones de z son constantes con respecto a y y viceversa. Al cambiar de coordenadas rectangulares, representadas por x y y, a coordenadas polares r y θ, en donde $y = r\cos\theta$ y $z = r\,\text{sen}\,\theta$. Esto es:

$$y^2 + z^2 = r^2\cos^2\theta + r^2\text{sen}^2\theta = r^2,$$

y el elemento de área $dydz$, en coordenadas rectangulares se reemplaza por $rdrd\theta$ en coordenadas polares. Dado que los límites $(-\infty, \infty)$ tanto para y como para z generan el plano completo yz, el plano correspondiente a r y a θ se genera mediante el empleo de los límites $(0, 2\pi)$ para θ y $(0, \infty)$ para r. De esta forma se tiene:

$$I^2 = \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} \exp(-r^2/2)rdrd\theta$$

$$= \frac{1}{2\pi}\int_0^{2\pi} d\theta \int_0^{\infty} \exp(-r^2/2)rdr$$

$$= \frac{\theta}{2\pi}\Bigg|_0^{2\pi} \cdot [-\exp(-r^2/2)]\Bigg|_0^{\infty}$$

$$= 1,$$

y, por lo tanto, (5.1) es una función de densidad de probabilidad.

APÉNDICE

Demostración del teorema 5.1

La demostración que aquí se presenta se basa en el hecho de que una función generadora de momentos define, de manera única, a una distribución. Se demostrará que la función generadora de momentos de Y tiende a una distribución normal conforme $n \to \infty$. X es una variable aleatoria binomial:

$$m_X(t) = [(1 - p) + pe^t]^n.$$

Entonces:

$$m_Y(t) = E(e^{tY}) = E\left\{\exp\left[t(X - np)/\sqrt{np(1 - p)}\right]\right\}$$

$$= \exp\left[-npt/\sqrt{np(1 - p)}\right] E\left\{\exp\left[tX/\sqrt{np(1 - p)}\right]\right\},$$

donde $E\left\{\exp\left[tX/\sqrt{np(1 - p)}\right]\right\}$ es la función generadora de momentos de X con argumento $t/\sqrt{np(1 - p)}$. De esta forma se tiene:

$$m_Y(t) = \exp\left[-npt/\sqrt{np(1 - p)}\right]\left\{(1 - p) + p\exp\left[t/\sqrt{np(1 - p)}\right]\right\}^n;$$

pero:

$$\exp\left[-npt/\sqrt{np(1 - p)}\right] = \left\{\exp\left[-pt/\sqrt{np(1 - p)}\right]\right\}^n$$

y:

$$m_Y(t) = \left\{(1 - p)\exp\left[-pt/\sqrt{np(1 - p)}\right] + p\exp\left[\frac{t}{\sqrt{np(1 - p)}} - \frac{pt}{\sqrt{np(1 - p)}}\right]\right\}^n$$

$$= \left\{(1 - p)\exp\left[-pt/\sqrt{np(1 - p)}\right] + p\exp\left[(1 - p)t/\sqrt{np(1 - p)}\right]\right\}^n.$$

En la última expresión, al expander ambas funciones exponenciales en una serie de potencias, se tiene:

$$(1 - p)\exp\left[-pt/\sqrt{np(1-p)}\right] = (1-p) - \frac{(1-p)pt}{\sqrt{np(1-p)}} + \frac{(1-p)p^2t^2}{2np(1-p)}$$

$$+ \text{ términos en } (-1)^k\left(\frac{1}{n}\right)^{k/2}, \qquad k = 3, 4, \ldots$$

$$= (1-p) - \frac{(1-p)pt}{\sqrt{np(1-p)}} + \frac{pt^2}{2n}$$

$$+ \text{ términos en } (-1)^k\left(\frac{1}{n}\right)^{k/2}, \qquad k = 3, 4, \ldots$$

y:

$$p\exp\left[(1-p)t/\sqrt{np(1-p)}\right] = p + \frac{(1-p)pt}{\sqrt{np(1-p)}} + \frac{(1-p)^2pt^2}{2np(1-p)}$$

$$+ \text{ términos en } \left(\frac{1}{n}\right)^{k/2}, \qquad k = 3, 4, \ldots$$

$$= p + \frac{(1-p)pt}{\sqrt{np(1-p)}} + \frac{(1-p)t^2}{2n}$$

$$+ \text{ términos en } \left(\frac{1}{n}\right)^{k/2}, \qquad k = 3, 4, \ldots .$$

Al sustituir los resultados anteriores en $m_Y(t)$ y agrupar términos,

$$m_Y(t) = \left[1 + \frac{t^2}{2n} + \text{términos en}\left(\frac{1}{n}\right)^{k/2}\right]^n, \qquad k = 3, 4, \ldots .$$

Dado que todos los términos que contienen a $(1/n)^{k/2}$, $k = 3, 4, \ldots$, tienen exponentes mayores que uno, puede factorizarse el término $1/n$. De esta forma se tiene que:

$$m_Y(t) = \left\{1 + \frac{1}{n}\left[\frac{t^2}{2} + \text{términos en}\left(\frac{1}{n}\right)^{(k-2)/2}\right]\right\}^n, \qquad k = 3, 4, \ldots .$$

Por definición:

$$\lim_{n\to\infty}\left(1 + \frac{u}{n}\right)^n = e^u;$$

entonces, conforme $n \to \infty$, la última expresión para $m_Y(t)$ es idéntica a esta forma, con u representando a todo lo que se encuentra entre paréntesis de esta expresión. Pero conforme $n \to \infty$, todos los términos de u, excepto el primero, tienen un valor

de cero, dado que todos tienen potencias positivas de n en sus denominadores. De acuerdo con lo anterior.

$$\lim_{n \to \infty} m_Y(t) = \exp(t^2/2),$$

que es la función generadora de momentos de la distribución normal estándar.

CAPÍTULO SEIS

Distribuciones conjuntas de probabilidad

6.1 Introducción

En los capítulos anteriores se consideraron conceptos probabilísticos tomando en cuenta una variable aleatoria a la vez. Sin embargo, muchas veces resulta de interés medir más de una característica de algún fenómeno aleatorio. Por ejemplo, en un proceso de producción en el que se tiene determinado número de artículos producidos en un tiempo definido, es muy común que el interés no sólo recaiga en el número de artículos que se encuentran listos para su venta inmediatamente después de su fabricación, sino también en el número que, después de reprocesarse, cae en la categoría anterior o en el número de artículos que serán desechados. Otro ejemplo puede ser que, al estudiar la contaminación del agua en general, se mida la concentración de varios contaminantes presentes en ésta. De los ejemplos anteriores surge la necesidad de estudiar modelos de probabilidad que contengan más de una variable aleatoria. Estos modelos reciben el nombre de *modelos multivariados,* mientras que los modelos con una sola variable reciben el nombre de *univariados.* En este capítulo se examinarán conceptos generales para distribuciones de probabilidad discretas y continuas con dos variables aleatorias. La extensión de estos conceptos a un mayor número de variables aleatorias resulta directa.

6.2 Distribuciones de probabilidad bivariadas

En esta sección se considerarán las definiciones pertinentes para distribuciones, tanto discretas como continuas, de dos variables aleatorias.

Definición 6.1 Sean X y Y dos variables aleatorias discretas. La probabilidad de que $X = x$ y $Y = y$ está determinada por la función de probabilidad bivariada

$$p(x, y) = P(X = x, Y = y),$$

en donde $p(x, y) \geqslant 0$ para toda x, y, de X, Y, y $\Sigma_x \Sigma_y p(x, y) = 1$. La suma se efectúa sobre todos los valores posibles de x y y.

Con base en la definición 6.1, la función de distribución acumulativa bivariada es la probabilidad conjunta de que $X \leqslant x$ y $Y \leqslant y$, dada por

$$F(x, y) = P(X \leqslant x, Y \leqslant y) = \sum_{x_i \leqslant x} \sum_{y_i \leqslant y} p(x_i, y_i). \tag{6.1}$$

La expresión anterior es una extensión del caso univariado. La función de probabilidad conjunta de dos variables aleatorias da origen a las probabilidades puntuales conjuntas, y la función de distribución bivariada es una función escalonada creciente para cada probabilidad puntual distinta de cero, de manera tal que $X = x$ y $Y = y$.

Ejemplo 6.1 Con base en la experiencia se sabe que la proporción de unidades útiles producidas por un proceso de manufactura es p_1, y las proporciones de unidades enviadas a reprocesar y desechadas, son p_2 y P_3, respectivamente. Si se supone que el número de unidades que se produce en un lapso dado es n y que además éstas constituyen un conjunto de ensayos independientes de manera que $p_1 + p_2 + p_3 = 1$, desarrollar una expresión para la probabilidad de tener, de manera exacta, x_1, x_2 y x_3 unidades útiles, reprocesables y desechadas, respectivamente.

Lo que se pide es una extensión de la distribución binomial univariada. A pesar de que existen tres resultados mutuamente excluyentes (útil, reprocesable y desechado), sólo es necesario definir dos variables aleatorias dado que, para cualquier número específico de cada una, la suma de las tres es n. Por consiguiente, sean X y Y las variables aleatorias que representan el número de unidades útiles y reprocesables, respectivamente, del total de unidades n. De esta manera, si $X = x$ y $Y = y$, entonces el número de unidades que deben desecharse es $n-x-y$. Por la hipótesis de independencia, la probabilidad de tener una secuencia específica de resultados es

$$p_1^x p_2^y (1 - p_1 - p_2)^{n-x-y}.$$

Dado que existen $n!/[x!y!(n - x - y)!]$ formas igualmente probables para que ocurra una secuencia de resultados específica, la probabilidad conjunta de tener, de manera exacta, x, y, y $n-x-y$ unidades útiles, reprocesables y desechadas, respectivamente, es

$$p(x, y; n, p_1, p_2) = \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y (1 - p_1 - p_2)^{n-x-y},$$
$$x, y = 0, 1, 2, \ldots, n, \tag{6.2}$$

en donde $p_3 = 1 - p_1 - p_2$. La expresión (6.2) es la función de probabilidad conjunta de lo que se conoce como la distribución trinomial. Los parámetros de esta distribución son n, p_1 y p_2, dado que p_3 se determina de manera exacta si se conocen

p_1 y p_2. La distribución trinomial se ha aplicado, de manera extensa, a situaciones en que existen tres resultados distintos, como en las encuestas sobre la preferencia del consumidor en relación a tres marcas comerciales o en encuestas de tipo político en que se pide la opinión con respecto a tres candidatos.

Si existen k resultados distintos excluyentes con probabilidades $p_1, p_2, \dots p_k$, respectivamente, entonces para n ensayos independientes, la distribución trinomial se generaliza para originar la distribución multinomial cuya función de probabilidad es:

$$p(x_1, x_2, \dots, x_{k-1}; n, p_1, p_2, \dots, p_{k-1}) = \frac{n!}{x_1!x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

$$x_i = 0, 1, 2, \dots, n \quad \text{for } i = 1, 2, \dots, k, \qquad (6.3)$$

en donde $x_k = n - x_1 - x_2 - \cdots - x_{k-1}$ y $p_k = 1 - p_1 - p_2 - \cdots - p_{k-1}$.

Definición 6.2 Sean X y Y dos variables aleatorias continuas. Si existe una función $f(x, y)$ tal que la probabilidad conjunta:

$$P(a < X < b, c < Y < d) = \int_a^b \int_c^d f(x, y) dy dx$$

para cualquier valor de a, b, c, y d en donde $f(x, y) \geq 0$, $-\infty < x, y < \infty$, y $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dydx = 1$, entonces $f(x, y)$ es la función de densidad de probabilidad bivariada de X y Y.

La función de densidad de probabilidad de dos variables aleatorias continuas X y Y es una superficie en el espacio de tres dimensiones donde el volumen por debajo de ésta y por encima de un rectángulo específico $a < X < b$ y $c < Y < d$ es igual a la probabilidad de que las variables aleatorias tomen valores iguales a los puntos que se encuentren dentro del rectángulo.

La función de distribución bivariada acumulativa de X y Y es la probabilidad conjunta de que $X \leq x$ y $Y \leq y$, dada por:

$$P(X \leq x, Y \leq y) = F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) dv du. \qquad (6.4)$$

Por lo tanto, la función de densidad bivariada se encuentra diferenciando $F(x, y)$ con respecto a x y y; es decir,

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}. \qquad (6.5)$$

Ejemplo 6.2 Sean X y Y dos variables aleatorias continuas con función de densidad de probabilidad conjunta dada por:

$$f(x, y) = \begin{cases} (x + y) & 0 \leq x, y \leq 1, \\ 0 & \text{para cualquier otro valor} \end{cases}$$

Graficar la función de densidad de probabilidad conjunta, determinar la función de distribución acumulativa conjunta y obtener la probabilidad conjunta de que $X \leq 1/2$ y $Y \leq 3/4$.

La gráfica de la función de densidad conjunta se ilustra en la figura 6.1. Nótese que $f(x, y)$ es una función de densidad de probabilidad conjunta, dado que

$$\int_0^1\int_0^1 (x + y)dydx = \int_0^1 \left(xy + \frac{y^2}{2}\right)\bigg|_0^1 dx = \int_0^1 \left(x + \frac{1}{2}\right)dx = 1.$$

Entonces

$$F(x, y) = \int_0^x\int_0^y (u + v)dvdu = \int_0^x \left(uy + \frac{y^2}{2}\right)du = xy(x + y)/2, \qquad 0 \leq x, y \leq 1.$$

De esta forma se tiene

$$F(1/2, 3/4) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{3}{4}\right)\left(\frac{1}{2} + \frac{3}{4}\right) = 15/64.$$

Además

$$\frac{\partial F(x, y)}{\partial x} = xy + \frac{y^2}{2}$$

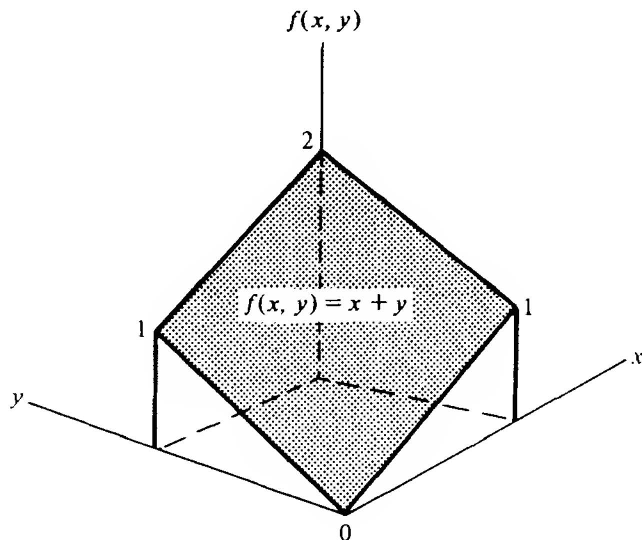


FIGURA 6.1 Gráfica de la función de densidad conjunta $f(x, y) = x + y$

y

$$\frac{\partial^2 F(x, y)}{\partial x\, \partial y} = x + y = f(x, y).$$

6.3 Distribuciones marginales de probabilidad

Es posible determinar varias distribuciones marginales para cualquier distribución de probabilidad que contenga más de dos variables aleatorias. Por ejemplo, si X y Y son variables aleatorias discretas, la suma de la función de probabilidad bivariada sobre todos los valores posibles de Y dará origen a la función de probabilidad univariada de X. Por otro lado, si X y Y son variables aleatorias continuas, la integración de la función de densidad de probabilidad bivariada sobre el intervalo completo de variación de Y generará la función de densidad de probabilidad univariada de X. De acuerdo con lo anterior, se formulan las siguientes definiciones:

Definición 6.3 Sean X y Y dos variables aleatorias discretas con una función de probabilidad conjunta $p(x, y)$. Las funciones marginales de probabilidad de X y de Y están dadas por

$$p_X(x) = \sum_y p(x, y)$$

y

$$p_Y(y) = \sum_x p(x, y),$$

respectivamente.

Definición 6.4 Sean X y Y dos variables aleatorias continuas con una función de densidad de probabilidad conjunta $f(x, y)$. Las funciones de densidad de probabilidad de X y de Y están dadas por

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

y

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx,$$

respectivamente.

Para variables aleatorias continuas conjuntas, si se conoce la función de distribución acumulativa $F(x, y)$, las distribuciones acumulativas marginales de X y Y se obtienen de la siguiente forma:

$$P(X \leq x) = F_X(x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(t, y)dydt,$$

y

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = F(x, \infty). \tag{6.6}$$

De manera similar

$$P(Y \leqslant y) = F_Y(y) = \int_{-\infty}^{y}\int_{-\infty}^{\infty} f(x, t)dxdt = \int_{-\infty}^{y} f_Y(t)dt = F(\infty, y). \tag{6.7}$$

Así puede determinarse la distribución acumulativa marginal de X dejando que Y tome un valor igual al límite superior de la función de distribución conjunta de X y Y.

Ejemplo 6.3 Sean X y Y dos variables aleatorias continuas con una función de densidad de probabilidad conjunta:

$$f(x, y) = \begin{cases} 3x(1 - xy) & 0 \leqslant x, y \leqslant 1, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

Obtener las distribuciones de densidad marginal y acumulativa de X y Y.

La función de densidad marginal de X es

$$f_X(x) = 3\int_0^1 x(1 - xy)dy = 3\left(xy - \frac{x^2y^2}{2}\right)\Bigg|_0^1 = 3x\left(1 - \frac{x}{2}\right).$$

De manera similar para Y

$$f_Y(y) = 3\int_0^1 x(1 - xy)dx = 3\left(\frac{x^2}{2} - \frac{x^3y}{3}\right)\Bigg|_0^1 = (3 - 2y)/2.$$

La distribución acumulativa conjunta de X y Y es

$$F(x, y) = 3\int_0^x\int_0^y u(1 - uv)dvdu = 3\int_0^x \left(uy - \frac{u^2y^2}{2}\right)du$$

$$= x^2y(3 - xy)/2, \qquad 0 \leqslant x, y \leqslant 1.$$

Por lo tanto, las distribuciones acumulativas marginales de X y Y están dadas por

$$F_X(x) = F(x, 1) = x^2(3 - x)/2, \qquad 0 \leqslant x \leqslant 1,$$

y

$$F_Y(y) = F(1, y) = y(3 - y)/2, \qquad 0 \leqslant y \leqslant 1,$$

respectivamente.

y

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = F(x, \infty). \tag{6.6}$$

De manera similar

$$P(Y \leq y) = F_Y(y) = \int_{-\infty}^{y}\int_{-\infty}^{\infty} f(x, t)dxdt = \int_{-\infty}^{y} f_Y(t)dt = F(\infty, y). \tag{6.7}$$

Así puede determinarse la distribución acumulativa marginal de X dejando que Y tome un valor igual al límite superior de la función de distribución conjunta de X y Y.

Ejemplo 6.3 Sean X y Y dos variables aleatorias continuas con una función de densidad de probabilidad conjunta:

$$f(x, y) = \begin{cases} 3x(1 - xy) & 0 \leq x, y \leq 1, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

Obtener las distribuciones de densidad marginal y acumulativa de X y Y.

La función de densidad marginal de X es

$$f_X(x) = 3\int_0^1 x(1 - xy)dy = 3\left(xy - \frac{x^2y^2}{2}\right)\Big|_0^1 = 3x\left(1 - \frac{x}{2}\right).$$

De manera similar para Y

$$f_Y(y) = 3\int_0^1 x(1 - xy)dx = 3\left(\frac{x^2}{2} - \frac{x^3y}{3}\right)\Big|_0^1 = (3 - 2y)/2.$$

La distribución acumulativa conjunta de X y Y es

$$F(x, y) = 3\int_0^x\int_0^y u(1 - uv)dvdu = 3\int_0^x \left(uy - \frac{u^2y^2}{2}\right)du$$

$$= x^2y(3 - xy)/2, \qquad 0 \leq x, y \leq 1.$$

Por lo tanto, las distribuciones acumulativas marginales de X y Y están dadas por

$$F_X(x) = F(x, 1) = x^2(3 - x)/2, \qquad 0 \leq x \leq 1,$$

y

$$F_Y(y) = F(1, y) = y(3 - y)/2, \qquad 0 \leq y \leq 1,$$

respectivamente.

6.4 Valores esperados y momentos para distribuciones bivariadas

En esta sección se tratarán los conceptos de valor esperado y momentos para distribuciones conjuntas de probabilidad.

Definición 6.5 Sean X y Y dos variables aleatorias que se distribuyen conjuntamente. El valor esperado de una función de X y de Y, $g(x, y)$, se define como

$$E[g(X, Y)] = \sum_x \sum_y g(x, y)p(x, y)$$

si X y Y son discretas, o

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)f(x, y)dydx$$

si X y Y son continuas, en donde $p(x, y)$ y $f(x, y)$ son las funciones de probabilidad y de densidad de probabilidad conjuntas, respectivamente.

Sin pérdida de generalidad, se restringirá la presentación al caso continuo. Como consecuencia de la definición 6.5, el r-ésimo momento de X alrededor del cero es

$$\begin{aligned} E(X^r) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^r f(x, y)dydx \\ &= \int_{-\infty}^{\infty} x^r f_X(x)dx. \end{aligned} \tag{6.8}$$

De manera similar

$$E(Y^r) = \int_{-\infty}^{\infty} y^r f_Y(y)dy. \tag{6.9}$$

El r y s-ésimo *momento producto* de X y Y alrededor del origen es:

$$E(X^r Y^s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^r y^s f(x, y)dydx, \tag{6.10}$$

y alrededor de las medias es

$$E\{(X - \mu_X)^r (Y - \mu_Y)^s\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)^r (y - \mu_Y)^s f(x, y)dydx, \tag{6.11}$$

en donde r y s son enteros, no negativos. Nótese que el r-ésimo momento de X alrededor del cero se obtiene de (6.10) con $s = 0$. De manera similar, el r-ésimo momento central de X puede determinarse a partir de (6.11) con $s = 0$.

De particular importancia es el momento producto alrededor de las medias cuando $r = s = 1$. Este momento producto recibe el nombre de *covarianza de* X y Y, y se

encuentra definido por

$$Cov(X, Y) = E\{(X - \mu_X)(Y - \mu_Y)\}. \tag{6.12}$$

Al igual que la varianza, que es una medida de la dispersión de una variable aleatoria, la covarianza es una medida de la variabilidad conjunta de X y de Y. De esta forma, la covarianza es una medida de asociación entre los valores de X y de Y y sus respectivas dispersiones. Si, por ejemplo, se tiene una alta probabilidad de que valores grandes de X se encuentren asociados con valores grandes de Y, la covarianza será positiva. Por otro lado, si existe una alta probabilidad de que valores grandes de X se encuentren asociados con valores pequeños de Y o viceversa, la covarianza será negativa. Se demostrará posteriormente que la covarianza es cero si X y Y son estadísticamente independientes.

Desarrollando el miembro derecho de (6.12) se tiene

$$\begin{aligned} E\{(X - \mu_X)(Y - \mu_Y)\} &= E[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \\ &= E(XY) - \mu_X\mu_Y; \end{aligned}$$

de esta forma

$$Cov(X, Y) = E(XY) - E(X)E(Y). \tag{6.13}$$

Si la covarianza de X y de Y se divide por el producto de las desviaciones estándar de X y de Y, el resultado es una cantidad sin dimensiones que recibe el nombre de *coeficiente de correlación* y que se denota por $\rho(X, Y)$:*

$$\rho(X, Y) = Cov(X, Y)/\sigma_X\sigma_Y. \tag{6.14}$$

Se puede demostrar que el coeficiente de correlación se encuentra contenido en el intervalo $-1 \leq \rho \leq 1$. De hecho ρ es la covarianza de dos variables aleatorias estandarizadas X' y Y' en donde $X' = (X - \mu_X)/\sigma_X$ y $Y' = (Y - \mu_Y)/\sigma_Y$. Esto significa que el coeficiente de correlación es sólo una medida estandarizada de la asociación lineal que existe entre las variables aleatorias X y Y en relación con sus dispersiones. El valor $\rho = 0$ indica la ausencia de cualquier asociación lineal, mientras que los valores -1 y $+1$ indican relaciones lineales perfectas negativa y positiva, respectivamente. En este punto es necesario señalar que debe rechazarse cualquier otra interpretación de la palabra "correlación". Después se expondrá con detalle el coeficiente de correlación cuando se estudie el análisis de regresión.

Ejemplo 6.4 Sean X y Y dos variables aleatorias con una función de densidad conjunta de probabilidad.

$$f(x, y) = \begin{cases} \frac{2}{3}(x + y)\exp(-x) & x > 0, \quad 0 < y < 1, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

* Se omitirá la identificación de las variables aleatorias cuando sea necesario.

Obtener la covarianza y el coeficiente de correlación de X y de Y.

Si se toman los valores esperados apropiados, se tiene

$$
\begin{aligned}
E(X) &= \frac{2}{3}\int_0^{\infty}\int_0^{1} (x^2 + xy)\exp(-x)dydx \\
&= \frac{2}{3}\int_0^{\infty} (x^2 + x/2)\exp(-x)dx \\
&= \frac{2}{3}\int_0^{\infty} x^2 \exp(-x)dx + \frac{1}{3}\int_0^{\infty} x\exp(-x)dx \\
&= \frac{2\Gamma(3)}{3} + \frac{\Gamma(2)}{3} \\
&= 5/3;
\end{aligned}
$$

$$
\begin{aligned}
E(X^2) &= \frac{2}{3}\int_0^{\infty}\int_0^{1} (x^3 + x^2y)\exp(-x)dydx \\
&= \frac{2}{3}\int_0^{\infty} x^3 \exp(-x)dx + \frac{1}{3}\int_0^{\infty} x^2\exp(-x)dx \\
&= \frac{2\Gamma(4)}{3} + \frac{\Gamma(3)}{3} \\
&= 14/3;
\end{aligned}
$$

$$
\begin{aligned}
E(Y) &= \frac{2}{3}\int_0^{\infty}\int_0^{1} (xy + y^2)\exp(-x)dydx \\
&= \frac{1}{3}\int_0^{\infty} x\exp(-x)dx + \frac{2}{9}\int_0^{\infty} \exp(-x)dx \\
&= \frac{\Gamma(2)}{3} + \frac{2}{9} \\
&= 5/9;
\end{aligned}
$$

$$
\begin{aligned}
E(Y^2) &= \frac{2}{3}\int_0^{\infty}\int_0^{1} (xy^2 + y^3)\exp(-x)dydx \\
&= \frac{2}{9}\int_0^{\infty} x\exp(-x)dx + \frac{1}{6}\int_0^{\infty} \exp(-x)dx \\
&= \frac{2\Gamma(2)}{9} + \frac{1}{6} \\
&= 7/18;
\end{aligned}
$$

$$E(XY) = \frac{2}{3}\int_0^\infty \int_0^1 (x^2y + xy^2)\exp(-x)dydx$$

$$= \frac{1}{3}\int_0^\infty x^2 \exp(-x)dx + \frac{2}{9}\int_0^\infty x \exp(-x)dx$$

$$= \frac{\Gamma(3)}{3} + \frac{2\Gamma(2)}{9}$$

$$= 8/9.$$

Por lo tanto

$$Cov(X, Y) = E(XY) - E(X)E(Y) = 8/9 - (5/3)(5/9) = -1/27.$$

Dado que

$$Var(X) = E(X^2) - E^2(X) = 17/9$$

y

$$Var(Y) = E(Y^2) - E^2(Y) = 13/162,$$

el coeficiente de correlación es

$$\rho(X, Y) = \frac{-1/27}{\sqrt{(17/9)(13/162)}} = -0.0951.$$

6.5 Variables aleatorias estadísticamente independientes

En el capítulo dos se mencionó que dos eventos son estadísticamente independientes si su probabilidad conjunta es igual al producto de sus probabilidades marginales. En esta sección se extenderá el concepto de independencia a variables aleatorias. A fin de asegurar la consistencia de la definición debe insistirse que para variables aleatorias estadísticamente independientes, la probabilidad conjunta $P(a < X < b, c < Y < d)$ es igual al producto de las probabilidades individuales $P(a < X < b)$ y $P(c < Y < d)$. En este punto se proporciona la siguiente definición:

Definición 6.6 Sean X y Y dos variables aleatorias con una distribución conjunta. Se dice que X y Y son estadísticas independientes si y sólo si,

$$p(x, y) = p_X(x)p_Y(y) \qquad \text{si } X \text{ y } Y \text{ son discretas}$$

o bien

$$f(x, y) = f_X(x)f_Y(y) \qquad \text{si } X \text{ y } Y \text{ son continuas,}$$

para toda x y y, en donde $p(x, y)$ y $f(x, y)$ son las funciones bivariadas de probabilidad y de densidad de probabilidad, respectivamente, y en donde $p_X(x)$, $p_Y(y)$, $f_X(x)$, y $f_Y(y)$ son las funciones de probabilidad marginal o de densidad de probabilidad marginal apropiadas.

Se desprende de esta definición que si X y Y son estadísticamente independientes, la probabilidad conjunta

$$\begin{aligned} P(a < X < b, c < Y < d) &= \int_a^b \int_c^d f(x, y)dydx \\ &= \int_a^b \int_c^d f_X(x)f_Y(y)dydx \\ &= \int_a^b f_X(x)dx \int_c^d f_Y(y)dy \\ &= P(a < X < b)P(c < Y < d). \end{aligned}$$

Para la misma condición,

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y)dydx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf_X(x)f_Y(y)dydx \\ &= \int_{-\infty}^{\infty} xf_X(x)dx \int_{-\infty}^{\infty} yf_Y(y)dy \\ &= E(X)E(Y). \end{aligned}$$

Si X y Y son estadísticamente independientes, entonces $Cov(X, Y) = \rho(X, Y) = 0$. Sin embargo debe hacerse hincapié en que la proposición inversa no es cierta. Es decir, una covarianza igual a cero no es una condición suficiente para asegurar la independencia entre variables aleatorias. Debe notarse que si X y Y no son estadísticamente independientes, son estadísticamente dependientes.

Se establecerán algunos resultados útiles con base en las definiciones 6.5 y 6.6. Sean X y Y dos variables aleatorias continuas con una función de densidad conjunta de probabilidad $f(x, y)$.

El valor esperado de una función lineal de X y Y es

$$\begin{aligned} E(aX + bY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (ax + by) f(x, y)dydx \\ &= a \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xf(x, y)dydx + b \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf(x, y)dydx \\ &= a\, E(X) + b\, E(Y) \end{aligned} \tag{6.15}$$

para cualquier valor de las constantes a y b.

La varianza de una función lineal de X y Y es

$$\begin{aligned} Var(aX + bY) &= E(aX + bY)^2 - E^2(aX + bY) \\ &= E(a^2X^2 + 2abXY + b^2Y^2) - [aE(X) + bE(Y)]^2 \end{aligned}$$

$$= a^2E(X^2) + 2abE(XY) + b^2E(Y^2)$$
$$- a^2E^2(X) - 2abE(X)E(Y) - b^2E^2(Y)$$
$$= a^2Var(X) + b^2Var(Y) + 2abCov(X, Y). \quad (6.16)$$

Como consecuencia de los resultados anteriores, se tiene que el valor esperado de la suma de X y Y es la suma de los correspondientes valores esperados de X y Y, y la varianza de la suma de X y Y es igual a la suma de las respectivas varianzas más la covarianza de X y Y. Además, si X y Y son estadísticamente independientes.

$$Var(aX + bY) = a^2Var(X) + b^2Var(Y). \quad (6.17)$$

La generalización de estos resultados a n variables aleatorias se hace por inducción y se establece en el siguiente teorema:

Teorema 6.1 Sean $X_1, X_2, \ldots, X_n$ n variables aleatorias con una función de densidad conjunta de probabilidad $f(x_1, x_2, \ldots, x_n)$. Entonces

$$E\left[\sum_{i=1}^{n} a_iX_i\right] = \sum_{i=1}^{n}\left[a_iE(X_i)\right]$$

$$Var\left[\sum_{i=1}^{n} a_iX_i\right] = \sum_{i=1}^{n} a_i^2Var(X_i) + \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} a_ia_jCov(X_i, X_j)$$

para cualquier constante a_i, $i = 1, 2, \ldots, n$.

Ejemplo 6.5 Un vendedor obtiene sus ingresos mediante la venta de dos productos distintos. Por experiencia sabe que el volumen de ventas de A no tiene ninguna influencia sobre el de B. Su ingreso mensual es el 10% del volumen, en dólares, del producto A y el 15% del volumen de B. Si en promedio las ventas del producto A ascienden a \$10 000 con una desviación estándar de \$2 000 y las de B a \$8 000 con una desviación estándar de \$1 000, obténgase el valor esperado y la desviación estándar del ingreso mensual del vendedor.

Sean X y Y dos variables aleatorias que representan el volumen de ventas en dólares de los productos A y B, respectivamente. Por hipótesis:

$$E(X) = 10\,000, \quad d.e.(X) = 2\,000; \quad E(Y) = 8\,000, \quad d.e.(Y) = 1\,000.$$

De esta forma se tiene

$$E(0.1X + 0.15Y) = 0.1\,E(X) + 0.15\,E(Y) = \$2\,200,$$

y

$$Var(0.1X + 0.15Y) = 0.01\,Var(X) + 0.0225\,Var(Y) = 62\,500.$$

La desviación estándar es de \$250.

6.6 Distribuciones de probabilidad condicional

Supóngase que un tanque de agua contiene dos contaminantes. Sean X y Y dos variables aleatorias que representan el nivel de estos contaminantes en una porción del tanque que a su vez se encuentra representada por una superficie rectangular. Supóngase que el nivel observado de concentración de Y es y, pero no se observa X. Si se conoce la función de densidad conjunta de probabilidad $f(x, y)$, se necesita obtener una función que proporcione la probabilidad de que el nivel de concentración de X esté contenido en un intervalo (a, b) dado el valor observado de Y. Considere la función

$$f(x, y)/f_Y(y),$$

en donde $f_Y(y)$ es la densidad marginal de Y. Si se mantiene constante a la variable aleatoria Y en el valor observado y de manera tal que $f_Y(y) > 0$. entonces $f(x, y)/f_Y(y)$ define una función no negativa de X cuya integral es 1, dado que por definición

$$\int_{-\infty}^{\infty} \frac{f(x, y)}{f_Y(y)} dx = \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f(x, y)dx = f_Y(y)/f_Y(y) = 1.$$

De esta forma, $f(x, y)/f_Y(y)$ es una función de densidad de probabilidad y la probabilidad de que $a < X < b$, dado que el nivel de concentración de Y es y, está dada por:

$$P(a < X < b \mid y) = \int_a^b \frac{f(x, y)}{f_Y(y)} dx. \tag{6.18}$$

Definición 6.7 Sean X y Y dos variables aleatorias con una función de densidad conjunta de probabilidad $f(x, y)$. La función de densidad de probabilidad condicional de la variable aleatoria X, denotada por $f(x \mid y)$, para un valor fijo y de Y, está definida por

$$f(x \mid y) = f(x, y)/f_Y(y),$$

en donde $f_Y(y)$ es la función de densidad de probabilidad de Y de manera tal que $f_Y(y) > 0$.

De manera análoga, la función de densidad de probabilidad condicional de Y para un valor fijo x de X se define como

$$f(y \mid x) = f(x, y)/f_X(x) \qquad f_X(x) > 0, \tag{6.19}$$

en donde $f_X(x)$ es la densidad marginal de X. Puede pensarse a $f(x \mid y)$ como una función que da la densidad de probabilidad a lo largo de una línea horizontal en el plano (x, y) correspondiente a un valor fijo y de Y. De manera similar, $f(y \mid x)$ es una función que da la densidad de probabilidad a lo largo de una línea vertical en el plano (x, y) correspondiente a un valor x de X.

Nótese que si la densidad condicional $f(x \mid y)$ por ejemplo, no contiene a y, entonces X es estadísticamente independiente de Y. Esto es, si X y Y son estadísticamente independientes, entonces

$$f(x, y) = f_X(x) f_Y(y)$$

y

$$\begin{aligned} f(x \mid y) &= f(x, y)/f_Y(y) \\ &= f_X(x) f_Y(y)/f_Y(y) \\ &= f_X(x). \end{aligned}$$

De manera similar, si

$$f(x, y) = f_X(x) f_Y(y),$$

entonces

$$\begin{aligned} f(y \mid x) &= f_X(x) f_Y(y)/f_X(x) \\ &= f_Y(y). \end{aligned}$$

Los valores esperados condicionales se definen de manera análoga a la señalada en la definición 6.5. Por ejemplo, los valores esperados condicionales de X puesto que $Y = y$, y de Y, ya que $X = x$, se definen como

$$E(X \mid y) = \int_{-\infty}^{\infty} x f(x \mid y) dx$$

y (6.20)

$$E(Y \mid x) = \int_{-\infty}^{\infty} y f(y \mid x) dy,$$

respectivamente. El valor esperado de X dado y es una función del punto fijo y y representa la media de X a lo largo de la línea correspondiente a y. Por simetría, el valor esperado condicional de Y dado x es una función de x y representa la media de Y a lo largo de la línea correspondiente a x. De manera similar,

$$Var(X \mid y) = E(X^2 \mid y) - E^2(X \mid y)$$

y (6.21)

$$Var(Y \mid x) = E(Y^2 \mid x) - E^2(Y \mid x),$$

en donde

$$E(X^2 \mid y) = \int_{-\infty}^{\infty} x^2 f(x \mid y) dx$$

y (6.22)

$$E(Y^2 \mid x) = \int_{-\infty}^{\infty} y^2 f(y \mid x) dy.$$

Ejemplo 6.6 Sean X y Y los niveles de concentración en ppm de dos contaminantes en una determinada porción de un tanque de agua. Si la función de densidad conjunta de probabilidad está dada por

$$f(x, y) = \begin{cases} (x + y)/8000 & 0 < x, y < 20, \\ 0 & \text{para cualquier otro valor,} \end{cases}$$

y si el nivel de concentración observado de Y es de 10 ppm, obtener la probabilidad de que el nivel de concentración de X sea, a lo más, 14 ppm. Obtener la media y la varianza condicional de X para $Y = 10$ ppm.

Dado que

$$f(x, y) = (x + y)/8000 \qquad 0 < x, y < 20,$$

se tiene

$$f_Y(y) = \frac{1}{8000}\int_0^{20} (x + y)dx = (y + 10)/400,$$

y la densidad de probabilidad condicional de X es

$$f(x \mid y) = (x + y)/20(y + 10),$$

la que se reduce a

$$f(x \mid Y = 10) = (x + 10)/400$$

para $Y = 10$. Por lo tanto,

$$\begin{aligned} P(X \leq 14 \mid Y = 10) &= \int_0^{14} f(x \mid Y = 10)dx \\ &= \frac{1}{400}\int_0^{14} (x + 10)dx \\ &= 0.595. \end{aligned}$$

Para la media y varianza condicional de X en $Y = 10$ se tiene

$$\begin{aligned} E(X \mid Y = 10) &= \int_0^{20} xf(x \mid Y = 10)dx \\ &= \frac{1}{400}\int_0^{20} (x^2 + 10x)dx \\ &= 11.67; \end{aligned}$$

$$E(X^2 \mid Y = 10) = \int_0^{20} x^2 f(x \mid Y = 10)dx$$

$$= \frac{1}{400}\int_0^{20} (x^3 + 10x^2)dx$$

$$= 166.67;$$

$$Var(X \mid Y = 10) = 30.56.$$

6.7 Análisis bayesiano: las distribuciones *a priori* y *a posteriori*

Se estableció en la sección 2.8 el teorema de Bayes para probabilidades condicionales de eventos discretos. En este contexto se examinará de manera breve cómo emplearlo para modificar el grado de creencia con respecto a los resultados de un fenómeno al tenerse nueva información de éste. Sin embargo, es más importante la representación que proporciona el teorema de Bayes para la distribución condicional de una variable aleatoria ya sea ésta cotinua o discreta. Tal representación es importante debido a que, como se verá en el capítulo 8, proporciona el mecanismo necesario sobre el cual se basa la inferencia bayesiana. En esta sección se examinarán los conceptos de distribución *a priori* y distribución *a posteriori* y se volverá a plantear el teorema de Bayes con estos conceptos.

Sea Y una variable aleatoria (discreta o continua) definida de manera tal que sus valores representan las posibles opciones en que puede ocurrir un fenómeno aleatorio antes de llevar a cabo un experimento. El grado de creencia del investigador con respecto a estas posibilidades se encuentra expresado por una función de probabilidad $p_Y(y)$, que recibe el nombre de *función de probabilidad a priori* de Y, si Y es discreta, o una función de densidad $f_Y(y)$, denominada *función de densidad de probabilidad a priori* de Y, si Y es continua. La especificación de la forma de $p_Y(y)$ o $f_Y(y)$ depende de la convicción del investigador con respecto a los valores de Y antes de que la información muestral se encuentre disponible. Esta convicción se puede basar en cualquier tipo de información que se encuentre disponible, incluyendo el juicio subjetivo. Sea $f(x \mid y)$ la función de densidad de probabilidad condicional de cualquier variable aleatoria X*, la cual representa evidencia muestral en función de una alternativa fija y de Y. La función $f(x \mid y)$ recibe el nombre de *función de verosimilitud* debido a que representa el grado de concordancia del resultado muestral x, dado el valor y de Y.

Cuando la información *a priori* con respecto a los valores de Y se combina con la información que proporcionó la muestra, el resultado es un conjunto de información modificada con respecto a la variable aleatoria Y. En otras palabras, la combinación de la distribución *a priori* y de la función de verosimilitud origina una distribución condicional para Y, dado el resultado muestral, que se conoce como la *distribución a posteriori* de Y. Esta combinación se hace de acuerdo con el teorema de Bayes, mismo que se replantea de la siguiente forma:

Teorema 6.2 Sea $p_Y(y)$ o $f_Y(y)$ la función de probabilidad o de densidad de probabilidad *a priori* de Y, respectivamente, y sea $f(x \mid y)$ la función de verosimilitud.

* Se supone que la variable aleatoria X es continua aunque también puede ser discreta.

Entonces la probabilidad *a posteriori* o función de densidad de probabilidad *a posteriori* de Y dada la evidencia muestral x, es

$$p(y \mid x) = \frac{f(x \mid y)p_Y(y)}{\sum_Y f(x \mid y)p_Y(y)} \quad \text{si } Y \text{ es discreta,} \tag{6.23}$$

$$f(y \mid x) = \frac{f(x \mid y)f_Y(y)}{\int_Y f(x \mid y)f_Y(y)dy} \quad \text{si } Y \text{ es continua.} \tag{6.24}$$

La función de probabilidad *a posteriori* $p(y \mid x)$ o la función de densidad de probabilidad *a posteriori* $f(y \mid x)$ reflejan el grado de creencia modificado del investigador con respecto a la variable aleatoria Y después de obtener información muestral. Dado que esta información se puede verificar de manera periódica, puede adoptarse fácilmente un punto de vista secuencial. En este contexto, la distribución *a posteriori* actual puede convertirse, en un futuro, en una distribución *a priori* cuando sea necesario llevar a cabo otra revisión con respecto a la variable aleatoria. La revisión periódica de las probabilidades se hace posible mediante el empleo sucesivo del teorema 6.2.

Es interesante notar que el denominador de (6.23) o (6.24) es la función de densidad de probabilidad marginal o no condicional de X; esto es,

$$f_X(x) = \sum_Y f(x \mid y)p_Y(y) \tag{6.25}$$

o

$$f_X(x) = \int_Y f(x \mid y)f_Y(y)dy, \tag{6.26}$$

dependiendo de cuando Y es discreta o continua, respectivamente. Además, el numerador de (6.23) o (6.24) es el producto de la función de verosimilitud y la función de probabilidad *a priori* y, de esta manera, es la probabilidad conjunta de X y Y expresada como

$$f(x, y) = f(x \mid y)p_Y(y) \quad \text{si } Y \text{ es discreta,} \tag{6.27}$$

o

$$f(x, y) = f(x \mid y)f_Y(y) \quad \text{si } Y \text{ es continua.} \tag{6.28}$$

Nótese que para (6.27) la función $f(x, y)$ es una mezcla bivariada de una variable aleatoria continua y otra discreta.

Ejemplo 6.7 Un vendedor de artículos domésticos nota que el número de personas que compran determinada marca de televisores varía aleatoriamente en el tiempo. El vendedor concluye que esta proporción es una variable aleatoria discreta que puede tomar los valores de 0.3, 0.35, 0.4 y 0.45, dependiendo de diversas consideraciones

de tipo económico. Con base en información previa, les asigna las probabilidades *a priori* 0.4, 0.3, 0.2 y 0.1, respectivamente. Una muestra de tamaño $n = 15$ revela que ocho de los televisores que se venden son de la marca de interés. Si se supone que para una proporción en particular p, el número de televisores de la marca que se vende para una muestra fija n es una variable aleatoria binomial, obtener las probabilidades *a posteriori*.

Sea X la variable aleatoria que representa el número de aparatos de la marca de interés que se venden de una muestra de tamaño n. El valor $X = 8$ para $n = 15$, representa la evidencia muestral condicionada sobre una proporción en particular p de preferencia del consumidor para esta marca. Por hipótesis X es binomial y su función de verosimilitud es

$$p(x; 15 \mid p) = \frac{15!}{(15 - x)!x!} p^x(1 - p)^{15-x}, \qquad x = 0, 1, 2, \ldots, 15.$$

Si $p = 0.3$, el valor de verosimilitud de la muestra es

$$P(X = 8 \mid p = 0.3) = p(8; 15 \mid 0.3) = \frac{15!}{(15 - 8)!8!}(0.3)^8(0.7)^{15-8} = 0.0348.$$

Para los demás valores de p se tiene

$$P(X = 8 \mid p = 0.35) = 0.071,$$
$$P(X = 8 \mid p = 0.4) = 0.1181,$$
$$P(X = 8 \mid p = 0.45) = 0.1647.$$

Nótese que las dos variables aleatorias son discretas. A pesar de lo anterior, puede emplearse el teorema de Bayes (6.23) para obtener las probabilidades *a posteriori*. La tabla 6.1 proporciona los detalles computacionales. La suma de las probabilidades tanto *a priori* como *a posteriori* debe ser igual a uno, dado que cada una de éstas es una distribución de probabilidad. En la figura 6.2 se ilustran las gráficas

TABLA 6.1 Determinación de las probabilidades *a posteriori* para el ejemplo 6.7

Valores de la proporción	*Probabilidad* a priori	*Verosimilitud de la muestra*	*Probabilidad a priori × verosimilitud*	*Probabilidad* a posteriori
0.3	0.4	0.0348	0.01392	0.01392/0.07531 = 0.1848
0.35	0.3	0.071	0.02130	0.02130/0.07531 = 0.2828
0.4	0.2	0.1181	0.02362	0.02362/0.07531 = 0.3137
0.45	0.1	0.1647	0.01647	0.01647/0.07531 = 0.2187
Totales	1.0		0.07531	1.0000

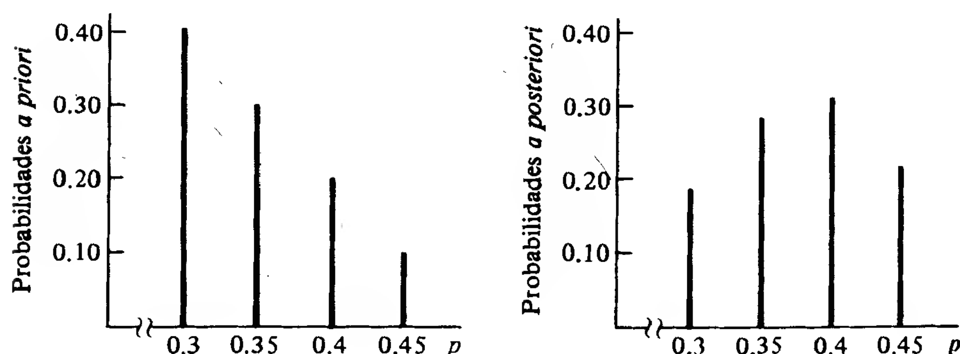


FIGURA 6.2 Probabilidades *a priori* y *a posteriori* para el ejemplo 6.7

de ambas distribuciones de probabilidad, las que muestran un desarrollo notable en las probabilidades para los cuatro valores posibles de p. También existe un desarrollo en los valores esperados de la preferencia del consumidor para esta marca. El valor esperado *a priori* es 0.35 y el valor esperado *a posteriori* es 0.3783.

Se mencionó en la sección 4.5 que la distribución binomial negativa es una alternativa adecuada del modelo de Poisson cuando la frecuencia de ocurrencia no es constante sobre el tiempo o el espacio. Por ejemplo, en las estadísticas de accidentes es poco probable que la frecuencia con que éstos se presentan entre grupos distintos sea constante e independiente sobre un lapso fijo. Lo anterior tiene como consecuencia que el punto de vista bayesiano sea una forma de análisis de estos datos mucho más apropiada.

Supóngase que todas las posibles frecuencias de ocurrencia pueden considerarse como valores de una variable aleatoria continua Λ, cuya distribución *a priori* es una distribución gama con una función de densidad dada por

$$f(\lambda; k, \theta) = \frac{1}{\Gamma(k)\theta^k} \lambda^{k-1}\exp(-\lambda/\theta), \qquad \lambda > 0. \tag{6.29}$$

Sea X una variable aleatoria que representa el número de accidentes que se observan en un grupo específico. Entonces puede argumentarse que X es una variable aleatoria de Poisson que depende de una λ específica de Λ, con una función de verosimilitud dada por

$$p(x \mid \lambda) = \lambda^x\exp(-\lambda)/x! \qquad x = 0, 1, 2, \ldots . \tag{6.30}$$

Antes de obtener la distribución *a posteriori* de Λ, se demostrará que la función de probabilidad marginal de X es la binomial negativa. Esto es, si para cada valor λ de Λ, X tiene una distribución de Poisson, entonces la distribución no condicional de X sobre todos los posibles valores de λ es la binomial negativa.

De (6.26) se desprende que la función de probabilidad marginal de X es:

$$p_X(x) = \int_0^\infty p(x \mid \lambda) f_\Lambda(\lambda) d\lambda. \tag{6.31}$$

Nótese que el integrando de (6.31) es la función de densidad conjunta de probabilidad de X y Λ, lo que da como resultado una mezcla bivariada de una variable aleatoria discreta con una continua.

La sustitución de (6.29) y (6.30) en (6.31) conduce a:

$$p_X(x) = \frac{1}{\Gamma(k)\theta^k x!} \int_0^\infty \lambda^{x+k-1} \exp\left[-\lambda\left(\frac{\theta+1}{\theta}\right)\right] d\lambda. \tag{6.32}$$

En el integrando de (6.32) sea $u = \lambda\ [(\theta + 1)/\theta]$; de esta forma $\lambda = [\theta/(\theta + 1)]u$ y $d\lambda = [\theta/(\theta + 1)]du$. Entonces

$$p_X(x) = \frac{1}{x!\Gamma(k)\theta^k} \int_0^\infty \theta/(\theta + 1)^{x+k} u^{x+k-1} \exp(-u) du$$

$$= \frac{\theta/(\theta + 1)^{x+k}\Gamma(x + k)}{x!\Gamma(k)\theta^k}$$

$$= \frac{\Gamma(x + k)}{x!\Gamma(k)} \left(\frac{1}{\theta + 1}\right)^k \left(\frac{\theta}{\theta + 1}\right)^x, \quad \begin{matrix} x = 0, 1, 2, \ldots, \\ k, \theta > 0. \end{matrix} \tag{6.33}$$

La expresión (6.33) es idéntica a la dada por (4.35), que es la función de probabilidad de la distribución binomial negativa para $k > 0$. Nótese que en (6.33), $p = 1/(\theta + 1)$ y $1 - p = \theta/(\theta + 1)$, de forma tal que $0 < p < 1$ dado que $\theta > 0$. Además, de (4.39) la media de X es

$$E(X) = \frac{k\theta/(\theta + 1)}{1/(\theta + 1)} = k\theta = E(\Lambda).$$

De esta manera, la distribución binomial negativa es una combinación de distribuciones de Poisson donde la frecuencia aleatoria de ocurrencia tiene una distribución gama cuya media es igual a la media de Poisson. Por esta razón la distribución binomial negativa también se conoce como una distribución compuesta de Poisson.

Mediante el empleo del teorema 6.2 y, en particular, de la expresión (6.24), se puede obtener la densidad de probabilidad *a posteriori* de Λ condicionada al resultado muestral x de la siguiente forma:

$$f(\lambda \mid x) = \frac{\lambda^{x+k-1}\exp\{-[(\theta + 1)/\theta]\lambda\}}{\Gamma(k)\theta^k x!} \Big/ \frac{\Gamma(x + k)}{x!\Gamma(k)} \left(\frac{1}{\theta + 1}\right)^k \left(\frac{\theta}{\theta + 1}\right)^x$$

$$= \frac{\lambda^{x+k-1}\exp\{-[(\theta + 1)/\theta]\lambda\}}{\Gamma(k)x!\theta^k} \cdot \frac{\Gamma(k)x!(\theta + 1)^{x+k}}{\Gamma(x + k)\theta^x}$$

$$= \frac{[(\theta + 1)/\theta]^{x+k}\lambda^{x+k-1}\exp\{-[(\theta + 1)/\theta]\lambda\}}{\Gamma(x + k)}, \quad \lambda > 0. \tag{6.34}$$

La comparación de (6.34) con la función de densidad de probabilidad de la distribución gama, dada por (5.45), muestra que la distribución *a posteriori* de Λ es una distribución gama con parámetros de forma $x + k$ y de escala $\theta/(\theta + 1)$. Debe notarse que si las distribuciones *a priori* y *a posteriori* pertenecen a la misma familia de distribuciones, como en el presente caso, ésta recibe el nombre de familia conjugada con respecto a la distribución de la muestra de datos. En este caso, la familia gama se conjuga con respecto a la distribución de Poisson.

Ejemplo 6.8 Supóngase que para las estadísticas de accidentes se decide asignar a la frecuencia de ocurrencia una distribución *a priori* gama con parámetro de forma dos y de escala tres. Supóngase que posteriormente se observan dos accidentes para una frecuencia en particular. Obtener la función de densidad *a posteriori* de la frecuencia, dado el resultado muestral, y compararla con la densidad *a priori*.

Sea Λ la frecuencia de ocurrencia. De (5.45) la densidad *a priori* de Λ es

$$f_\Lambda(\lambda; 2, 3) = \frac{1}{9}\lambda \exp(-\lambda/3), \qquad \lambda > 0.$$

Dado un resultado muestral $X = 2$, la densidad *a posteriori* de Λ que se obtiene de (6.34) es

$$f(\lambda; 4, 3/4 \mid x) = \frac{1}{6}(4/3)^4\lambda^3 \exp\left(-\frac{4}{3}\lambda\right), \qquad \lambda > 0.$$

En la figura 6.3 se proporciona una comparación entre las funciones de densidad *a priori* y *a posteriori*. De ésta es evidente que la densidad *a posteriori* se encuentra menos asimétrica que la densidad *a priori*. Nótese que la frecuencia media *a priori* es seis mientras que ésta misma *a posteriori* es tres.

En la sección 5.4 se mencionó que la distribución beta tiene un papel muy importante en la estadística bayesiana. Para ilustrar lo anterior considérese de nuevo el análisis bayesiano del parárametro de proporción de la distribución binomial.

Ejemplo 6.9 En un proceso de manufactura, el interés se centra alrededor de la proporción de artículos defectuosos. Dado que es poco probable que el proceso tenga cambios menores en un lapso determinado como distintos desarrollos, variaciones en la materia prima y otros que pueden influir en la proporción de artículos defectuosos, es razonable pensar la proporción de éstos como una variable aleatoria cuyos posibles valores se encuentran en el intervalo $(0, 1)$. Para una proporción dada de artículos defectuosos p, se sabe que el número x de éstos que se observa en una muestra aleatoria fija de n artículos es binomial. Esto es, la función de probabilidad condicional de X para n fijo, dado p, es

$$p(x; n \mid p) = \frac{n!}{(n - x)!x!} p^x(1 - p)^{n-x}, \qquad x = 0, 1, 2, \ldots n.$$

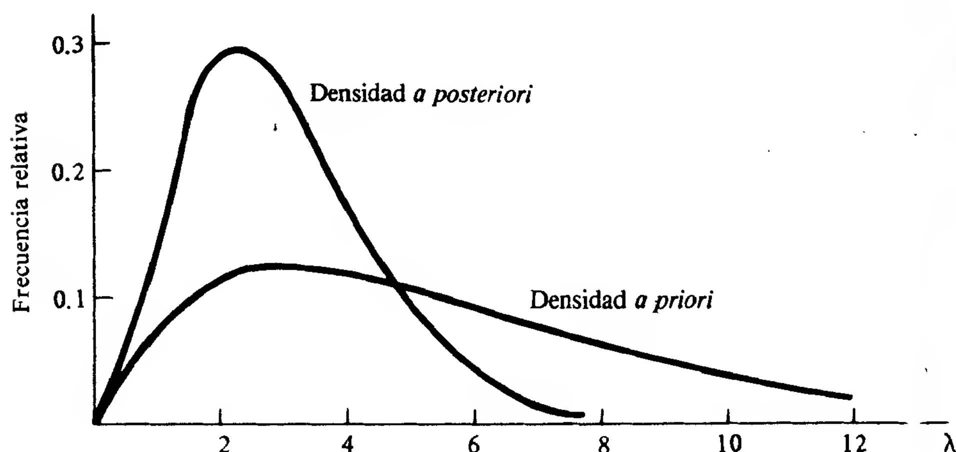


FIGURA 6.3 Densidades *a priori* y *a posteriori* para el ejemplo 6.8

Si la distribución *a priori* de la proporción de artículos defectuosos es una distribución beta con una función de densidad de probabilidad

$$f_P(p; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, p^{\alpha-1}(1 - p)^{\beta-1} \qquad 0 \leq p \leq 1, \tag{6.35}$$

demostrar que la distribución *a posteriori* de la proporción de artículos defectuosos, dado el número x de éstos, también es una distribución beta.

De (6.24) la densidad de probabilidad *a posteriori* de la proporción de artículos defectuosos es:

$$f(p \mid x) = \frac{p(x; n \mid p) f_P(p; \alpha, \beta)}{\displaystyle\int_0^1 p(x; n \mid p) f_P(p; \alpha, \beta)\,dp}$$

$$= \frac{\dfrac{n!}{(n-x)!\,x!}\, p^x(1 - p)^{n-x}\, \dfrac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, p^{\alpha-1}(1 - p)^{\beta-1}}{\dfrac{n!}{(n-x)!\,x!} \cdot \dfrac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \displaystyle\int_0^1 p^{x+\alpha-1}(1 - p)^{n+\beta-x-1}\,dp}$$

$$= \frac{p^{x+\alpha-1}(1 - p)^{n+\beta-x-1}}{\displaystyle\int_0^1 p^{x+\alpha-1}(1 - p)^{n+\beta-x-1}\,dp};$$

pero de (5.33), la integral $\int_0^1 p^{x+\alpha-1}(1-p)^{n+\beta-x-1}dp = B(x+\alpha, n+\beta-x)$. Por lo tanto, la densidad *a posteriori* es:

$$f(p \mid x) = \frac{p^{x+\alpha-1}(1-p)^{n+\beta-x-1}}{B(x+\alpha, n+\beta-x)}$$

$$= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(x+\alpha)\,\Gamma(n+\beta-x)}\, p^{x+\alpha-1}(1-p)^{n+\beta-x-1} \qquad 0 \leq p \leq 1, \qquad (6.36)$$

que es una densidad beta con parámetros $(x+\alpha)$ y $(n+\beta-x)$. Por lo tanto, la familia conjugada para la distribución binomial es la familia de distribuciones beta.

6.8 La distribución normal bivariada

En el capítulo cinco se estudió la distribución normal de una variable aleatoria. El concepto de distribución normal puede extenderse para incluir variables aleatorias. En particular, la distribución normal bivariada se emplea de manera extensa para describir el comportamiento probabilístico de dos variables aleatorias.

Definición 6.8 Se dice que las variables aleatorias X y Y tienen una *distribución normal bivariada* si su función de densidad conjunta de probabilidad está dada por

$$f(x,y) = \frac{1}{2\pi\,\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\} \quad -\infty < x, y < \infty, \qquad (6.37)$$

en donde

$$\mu_X = E(X), \quad \mu_Y = E(Y), \quad \sigma_X^2 = Var(X), \quad \sigma_Y^2 = Var(Y),$$

y ρ es el coeficiente de correlación de X y Y, definido en la sección 6.4.

La figura 6.4 ilustra la función de densidad normal bivariada que es una superficie tridimensional con forma de campana. Cualquier corte a través de la superficie da origen a una curva de forma normal univariada, mientras que planos paralelos al plano xy interceptan la superficie en elipses que reciben el nombre de *contornos de probabilidad constante*.

Es interesante notar que, a pesar de que $\rho = 0$ es una condición necesaria de independencia, para la distribución normal bivariada también es una condición suficiente. Eso es, si $\rho = 0$, entonces

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y}\exp\left[-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - \frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]$$

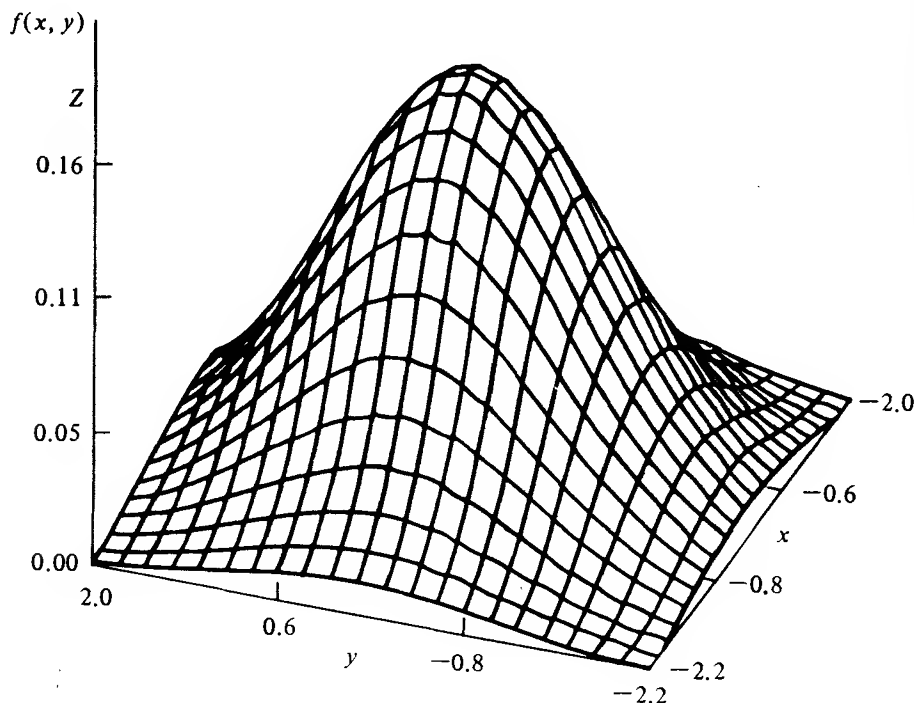


FIGURA 6.4 Densidad normal bivariada con $E(X) = E(Y) = 0$, $Var(X) = Var(Y) = 1$, y $\rho = 0$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_X} \exp\left[-(x - \mu_X)^2/2\sigma_X^2\right] \cdot \frac{1}{\sqrt{2\pi}\,\sigma_Y} \exp\left[-(y - \mu_Y)^2/2\sigma_Y^2\right]$$

$$= f_X(x) f_Y(y),$$

en donde $f_X(x)$ y $f_Y(y)$ son las densidades normales univariadas de X y Y, respectivamente.

Se puede demostrar que, mediante el empleo de (6.37) e integrando con respecto a y, la densidad marginal de X es normal con media μ_X varianza σ_X^2. De manera similar, la densidad marginal de Y es normal con media μ_Y y varianza σ_Y^2. Por la definición 6.7, la densidad de probabilidad condicional de X dado el valor y de Y es

$$f(x \mid y) = \frac{1}{\sqrt{2\pi\,\sigma_X^2(1 - \rho^2)}} \times \exp\left\{-\frac{1}{2\sigma_X^2(1 - \rho^2)}\left[x - \mu_X - \frac{\rho\,\sigma_X}{\sigma_Y}(y - \mu_Y)\right]^2\right\}. \quad (6.38)$$

La expresión (6.38) es una función de densidad de probabilidad normal con

$$E(X \mid y) = \mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y - \mu_Y) \quad \text{y} \quad Var(X \mid y) = \sigma_X^2(1 - \rho^2).$$

Se puede obtener una expresión similar para la densidad condicional de Y dado el valor x de X.

Ejemplo 6.10 Sean X y Y las desviaciones horizontal y vertical (sobre un plano), respectivamente, de un vehículo espacial tripulado con respecto al punto de aterrizaje de éste en el mar de la Tranquilidad. Supóngase que X y Y son dos variables aleatorias, independientes cada una, con una distribución normal bivariada y medias $\mu_X = \mu_Y = 0$ y varianzas iguales. ¿Cuál es la máxima desviación estándar permisible de X y Y, que cumplirá con el requisito de la NASA de tener una probabilidad de 0.99, de que el vehículo aterrice a no más de 500 ft del punto elegido, tanto en dirección vertical como horizontal?

Debido a la independencia y a la hipótesis de que $\sigma_X = \sigma_Y = \sigma$, la probabilidad conjunta es

$$\begin{aligned} P(-500 < X < 500,\ -500 < Y < 500) &= P(-500 < X < 500) \\ &\quad \cdot P(-500 < Y < 500) \\ &= P\left(-\frac{500}{\sigma} < Z < \frac{500}{\sigma}\right) \\ &\quad \cdot P\left(-\frac{500}{\sigma} < Z < \frac{500}{\sigma}\right) \\ &= P^2\left(-\frac{500}{\sigma} < Z < \frac{500}{\sigma}\right). \end{aligned}$$

Puesto que por hipótesis es

$$P^2\left(-\frac{500}{\sigma} < Z < \frac{500}{\sigma}\right) = 0.99,$$

$$P\left(-\frac{500}{\sigma} < Z < \frac{500}{\sigma}\right) = 0.99499$$

o

$$P\left(Z > \frac{500}{\sigma}\right) = 0.0025,$$

pero

$$P(Z > 2.81) = 0.0025;$$

por lo tanto $500/\sigma = 2.81$, y $\sigma_X = \sigma_Y \leq 177.94$ pies

Referencias

1. P. G. Hoel, *Introduction to mathematical statistics*, 4th ed., Wiley, New York, 1971.
2. R. V. Hogg and A. T. Craig, *Introduction to mathematical statistics*, 4th ed., Macmillan, New York, 1978.
3. B. W. Lindgren, *Statistical theory*, 3rd ed., Macmillan, New York, 1976.

Ejercicios

6.1. Se seleccionaron, aleatoriamente, 60 personas y se les preguntó su preferencia con respecto a tres marcas A, B y C. Éstas fueron de 27, 18 y 15 respectivamente. ¿Qué tan probable es este resultado si no existen otras marcas en el mercado y la preferencia se comparte por igual entre las tres?

6.2. Supóngase que de un proceso de producción se seleccionan, de manera aleatoria, 25 artículos. Este proceso de producción por lo general produce un 90% de artículos listos para venderse y un 7% reprocesables. ¿Cuál es la probabilidad de que 22 de los 25 artículos estén listos para venderse y que dos sean reprocesables?

6.3. Sean X y Y dos variables aleatorias continuas con una función de densidad conjunta de probabilidad dada por:

$$f(x, y) = \begin{cases} (3x - y)/5 & 1 < x < 2, \quad 1 < y < 3, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

a) Obtener la función de distribución conjunta acumulativa.
b) ¿Cuál es la probabilidad conjunta de que $X < 3/2$ y $Y < 2$?
c) Mediante el empleo de sus respuestas a la parte *a,* obtener las distribuciones acumulativas marginales de X y Y.
d) Obtener las funciones de densidad marginal de X y de Y.

6.4. Sean X y Y dos variables aleatorias continuas con una función de densidad conjunta de probabilidad dada por

$$f(x, y) = \begin{cases} x \exp[-x(y + 1)] & x, y > 0, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

a) Demostrar que $f(x, y)$ es una función de densidad conjunta de probabilidad.
b) ¿Cuál es la probabilidad conjunta de que $X < 2$ y $Y < 1$?
c) Obtener las funciones de densidad marginal de X y de Y.
d) ¿Son X y Y estadísticamente independientes?

6.5. Sean X y Y dos variables aleatorias discretas en donde los posibles valores que éstas pueden tomar son -1, 0, y 1. En la siguiente tabla se dan las probabilidades conjuntas para todos los posibles valores de X y Y.

		X = −1	X = 0	X = 1
Y	−1	1/16	3/16	1/16
	0	3/16	0	3/16
	1	1/16	3/16	1/16

a) Obtener las funciones de probabilidad marginal $p_X(x)$ y $p_Y(y)$.
b) ¿Las variables aleatorias X y Y son estadísticamente independientes?
c) Obtener $Cov(X, Y)$.

6.6. Para la función de densidad conjunta de probabilidad del ejercicio 6.3, obtener $Cov(X, Y)$ y $\rho(X, Y)$.

6.7. En función de su prioridad, un programa para computadora espera en la fila de entrada cierto tiempo, depués del cual lo ejecuta el procesador central en un lapso dado. La función de densidad conjunta para los tiempos de espera y ejecución se determina por

$$f(t_1, t_2) = \begin{cases} 2 \exp\left[-\left(\dfrac{t_1}{5} + 10t_2\right)\right] & t_1, t_2 > 0, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

Dada la distribución conjunta acumulativa:

$$F(t_1, t_2) = \begin{cases} [1 - \exp(-t_1/5)][1 - \exp(-10t_2)] & t_1, t_2 > 0, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

a) Obtener la probabilidad conjunta de que el tiempo de espera no sea mayor de ocho minutos y el de ejecución no sea mayor de 12 segundos.
b) Obtener las funciones de densidad marginal y deducir que estos lapsos son variables aleatorias independientes.

6.8. Las variables aleatorias X y Y representan las proporciones de los mercados correspondientes a dos productos distintos fabricados por la misma compañía y cuya función de densidad conjunta de probabilidad está dada por

$$f(x, y) = \begin{cases} (x + y) & 0 \leq x, y \leq 1, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

a) Obtener las funciones de densidad marginal de X y Y.
b) ¿Las variables aleatorias X y Y son estadísticamente independientes?
c) Si $X = 0.2$, obtener la función de densidad de probabilidad condicional de Y.

6.9. Las variables aleatorias X y Y representan el largo y ancho (en cm) de una hoja de acero. Si X y Y son independientes con funciones de densidad de probabilidad dadas por

$$f_X(x) = \begin{cases} 1, & 99 < x < 100, \\ 0 & \text{para cualquier otro valor.} \end{cases} \qquad f_Y(y) = \begin{cases} 1, & 49 < y < 50, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

úsese la definición de la varianza para obtener la varianza del área de la hoja de acero XY.

6.10. Sea X una variable aleatoria continua y Y discreta.
a) Si $f(x, y) = x^y \exp(-2x)/y!$, $x > 0$, $y = 0, 1, 2, \ldots$, obtener la función de probabilidad marginal de Y.
b) Obtener la función de probabilidad condicional de X para $Y = 2$.
c) Obtener $E(X \mid 2)$ y $Var(X \mid 2)$.

6.11. Sean X y Y dos variables aleatorias. Demostrar que $Var(aX - bY) = a^2\, Var(X) + b^2\, Var(Y) - 2ab\, Cov(X, Y)$, en donde a y b son constantes.

6.12. Sean X y Y dos variables aleatorias. Demostrar que $Cov(aX, bY) = ab\, Cov(X, Y)$, en donde a y b son constantes.

6.13. Si X y Y son dos variables aleatorias independientes $Var(X + Y) = Var(X - Y) = Var(X) + Var(Y)$. Comparar este resultado con $Var(X + Y)$ cuando $Var(X - Y)$ $Cov(X, Y) > 0$ o $Cov(X, Y) < 0$. ¿Qué puede concluirse?

6.14. Supóngase que la frecuencia Λ a la que ocurren accidentes automovilísticos en un lapso fijo es una variable aleatoria con una ditribución gama y parámetros de forma y escala igual a dos. Si para cada valor λ de Λ la distribución condicional del número de accidentes es una distribución de Poisson, obtener la función de probabilidad marginal de X y calcular las probabilidades para $X = 0, 1, 2 \ldots 10$. ¿Cómo son estas probabilidades al compararlas con las que se obtienen bajo la suposición de una frecuencia constante $\lambda = 4$?

6.15. Supóngase que la incidencia de cáncer pulmonar para un determinado número de personas adultas, sin importar sus hábitos de fumador, su edad, etc., es una variable aleatoria con distribución gama con parámetros de forma y escala iguales a dos. Para un grupo específico de personas, el número que presentarán cáncer pulmonar es una variable aleatoria de Poisson en donde el valor del parámetro de ésta depende de la incidencia de cáncer en este grupo. Obtener la probabilidad no condicional de que no más de dos personas desarrollen cáncer en este grupo.

6.16 En el ejercicio 6.15 supóngase que $x = 5$ adultos, de cierto número, desarrollarán cáncer. Obtener la densidad *a posteriori* de Λ dado x, calcular las medias y varianzas tanto *a priori* como *a posteriori* y comparar los resultados.

6.17 Supóngase que el gerente de una planta descubre que la proporción de artículos defectuosos en su proceso de producción no es constante sino que se comporta como una variable aleatoria. Sin ninguna evidencia, decide asignar una distribución beta con $\alpha = 1$ y $\beta = 24$ para la producción de artículos defectuosos.

a) Graficar la función de densidad *a priori* y obtener su media y su varianza.
b) Supóngase que el gerente toma una muestra $n = 12$ artículos y encuentra uno defectuoso. Bajo las hipótesis necesarias, obtener y graficar la función de densidad de probabilidad *a posteriori*.
c) Encontrar la media y la varianza *a posteriori* y compararlas con la media y la varianza *a priori*.
d) Hágase uso del ejercicio 5.24 para obtener la probabilidad *a posteriori* de que la proporción de artículos defectuosos sea a lo más 0.05.

6.18. Supóngase que la proporción de lanzamientos exitosos de satélites de comunicaciones es una variable aleatoria con distribución beta y parámetros $\alpha = 21$ y $\beta = 1$. Si de los últimos 12 lanzamientos uno ha fracasado, obtener la función de probabilidad *a posteriori* de la proporción de lanzamientos exitosos y calcular la probabilidad *a posteriori* para que la proporción de éstos sea mayor de 0.95. Emplée la expresión 5.44.

6.19. La función de densidad conjunta de probabilidad para la demanda mensual de dos productos es una distribución normal bivariada dada por

$$f(x, y) = \frac{1}{100\pi\sqrt{3}} \exp\left\{-\frac{2}{3}\left[\left(\frac{x-50}{10}\right)^2 - \left(\frac{x-50}{10}\right)\left(\frac{y-25}{10}\right) + \left(\frac{y-25}{10}\right)^2\right]\right\}.$$

a) ¿Cuál es el coeficiente de correlación entre X y Y?
b) ¿Cuál es la covarianza entre X y Y?
c) Obtener la función de densidad de probabilidad condicional $f(x \mid y)$.
d) Supóngase que la demanda de Y es 30. ¿Cuál es la probabilidad condicional de que X sea menor que 65?

6.20. Supóngase que el CI(X) y la calificación promedio de estudiantes no graduados de licenciatura Y son variables aleatorias que se encuentran distribuidas de manera conjunta como una distribución normal bivariada $\mu_Y = 100$, $\sigma_X = 10$, $\mu_Y = 3$, $\sigma_Y = 0.3$, y $Cov(X, Y) = 2.25$.

a) Si algún estudiante posee un CI de 120, ¿cuáles son los valores de la media y la desviación estándar condicionales para Y?
b) Dado que el CI es 120, obtener la probabilidad de que Y sea mayor de 3.5.
c) Supóngase que la calificación promedio de un estudiante es 2.8. ¿Cuál es la probabilidad de que esta persona tenga un CI mayor de 115?

CAPÍTULO SIETE

Muestras aleatorias y distribuciones de muestreo

7.1 Introducción

En el capítulo uno se mencionó que para comprender la esencia de la inferencia estadística es necesario comprender la naturaleza de una población y de una muestra. Una población representa el "estado de la naturaleza" o la forma de las cosas con respecto a un fenómeno aleatorio en particular, mismo que puede identificarse a través de una característica medible X. La manera en que ocurren las cosas en relación con X puede definirse por un modelo de probabilidad que recibe el nombre de distribución de probabilidad de la población. Por otro lado, una muestra es una colección de datos que se obtienen al llevar a cabo repetidos ensayos de un experimento para lograr una evidencia representativa acerca de la población en relación con la característica X. Si la manera de obtener la muestra es imparcial y técnicamente buena, entonces la muestra puede contener información útil con respecto al estado de la naturaleza y a partir de ello se podrán formular inferencias. Ahora bien, estas últimas son inductivas y, por lo tanto, están sujetas a riesgo, dado que representan un razonamiento que va de lo particular a lo general.

En los capítulos cuatro, cinco y seis se examinaron con detalle algunas distribuciones de probabilidad que pueden servir como modelo para la distribución de una población de interés. En los capítulos restantes el principal objetivo es examinar distintas técnicas por medio de las cuales puede aplicarse el proceso inductivo de la inferencia estadística para proporcionar resultados útiles y confiables. La inferencia estadística se define como *la colección de técnicas que permiten formular inferencias inductivas y que proporcionan una medida del riesgo de éstas.* En este capítulo se establecerán algunos conceptos teóricos básicos con respecto al muestreo y a la inferencia estadística. La aplicación de estos conceptos se dará con gran detalle en capítulos posteriores.

7.2 Muestras aleatorias

Como la inferencia estadística se formula con base en una muestra de objetos de la población de interés, el proceso por medio del cual se obtiene será aquél que asegure

la selección de una buena muestra. En el capítulo uno se expuso que una manera de obtener una buena muestra resulta cuando el proceso de muestreo proporciona, a cada objeto en la población, una oportunidad igual e independiente de ser incluido en la muestra. Si la población consiste de N objetos y de éstos se selecciona una muestra de tamaño n, el proceso de muestreo debe asegurar que cada muestra de tamaño n tenga la misma probabilidad de ser seleccionada. Este procedimiento conduce a lo que comúnmente se conoce como una *muestra aleatoria simple.* En este contexto, la palabra "aleatorio" sugiere una total imparcialidad en la selección de la muestra.

La naturaleza de la inferencia inductiva demanda una muestra aleatoria porque la selección de ésta se lleva a cabo con el fin de proporcionar los medios adecuados para que pueda formularse una inferencia con respecto a alguna característica de la población de interés. Por ejemplo, pueden formularse inferencias de ciertas condiciones que se suponen válidas para la población si la muestra que se observó se encuentra o no dentro de la variación muestral, misma que prevalecerá si las condiciones son verdaderas. De esta forma la calidad de la aleatoriedad en una muestra asegura la aplicación correcta de la probabilidad para evaluar el riesgo inherente en un proceso inductivo.

En este momento es importante estructurar el concepto de una muestra aleatoria simple empleando para ello los conceptos de probabilidad que se presentaron en los capítulos dos al seis. Para llevar a cabo lo anterior, primero se examinarán situaciones que se presentan, de manera frecuente, en los muestreos. La primera de éstas surge en muchos experimentos que involucran fenómenos aleatorios en la ingeniería y las ciencias físicas. En estos casos la población de interés no consiste en objetos tangibles a partir de los cuales se selecciona un cierto número para formar la muestra. Más bien, la población se considera constituida por un número infinito de posibles resultados para alguna característica medible de interés. Esta característica generalmente es una medición física como el nivel de concentración de un contaminante, la demanda de un producto o el tiempo de espera en un servicio. Sea X una característica medible y $f(x;\ \theta)$ la función de densidad de probabilidad de la distribución de la población. El siguiente procedimiento es una forma de muestreo para este tipo de población:

1. Se diseña un experimento y se lleva a cabo para proporcionar la observación X_1 de la característica medible X. El experimento se repite bajo las mismas condiciones proporcionando el valor X_2. El proceso se continúa hasta tener n observaciones $X_1, X_2, \ldots, X_n$ de la característica X.

En este procedimiento de muestreo, las observaciones muestrales se colectan a través de ensayos independientes que ocurren cada vez que el experimento se repite bajo condiciones idénticas para todos los factores que son controlables. En este contexto, cada observación del i-ésimo experimento se considera como una selección de la misma fuente que proporciona la observación de cualquier otro ensayo para X. En esencia, las observaciones bajo las mismas condiciones como resultado de repetidos ensayos independientes de un experimento, constituye lo que se denomina un *muestreo aleatorio con reemplazo.* De acuerdo con lo anterior, cada una de las observaciones $X_1, X_2, \ldots, X_n$ es una variable aleatoria cuya distribución de probabilidad es idéntica a la de la población.

Una situación diferente se presenta cuando se lleva a cabo una selección de objetos tangibles de una población que consiste en un número finito de objetos (seres humanos, animales, componentes mecánicos o eléctricos, etc.). La característica medible de interés puede ser un atributo, como el estado de un componente (defectuoso o no defectuoso), la opinión de una persona con respecto a cierto tema (a favor o en contra) o una medición cuantitativa como el CI de una persona o el tiempo de duración de un componente. Existen dos formas para obtener muestras aleatorias de este tipo de población:

2. Después de llevar a cabo una mezcla adecuada de los objetos de la población, se extrae uno y se observa la característica medible. Esta observación será X_1. El objeto se regresa a la población y ésta vuelve a mezclarse; después se extrae el segundo objeto. X_2 se constituye por la segunda observación. El proceso se continúa de esta forma hasta que se han extraído n objetos para tener una muestra de observaciones $X_1, X_2, \ldots, X_n$ de la característica X.
3. Después de una mezcla adecuada de los objetos que constituyen la población, n de éstos se seleccionan uno después de otro sin reemplazo. Este proceso proporciona una muestra de observaciones $X_1, X_2, \ldots, X_n$ de la característica X.

Nótese que la técnica 2 constituye un muestreo con reemplazo y la técnica 3 es un muestreo sin reemplazo. En el contexto general de una muestra aleatoria simple, la técnica recibe el nombre de aleatoria. Cuando los objetos se extraen después de una selección equitativa. Por consiguiente, la técnica de muestreo dos recibe el nombre de muestreo aleatorio con reemplazo, y la técnica tres el de muestreo aleatorio sin reemplazo. En la técnica dos, cada una de las observaciones $X_1, X_2, \ldots, X_n$ es una variable aleatoria cuya distribución de probabilidad es identica a la de la población, puesto que en cada extracción ésta tiene su forma original. En la técnica de muestreo tres, las observaciones $X_1, X_2 \ldots, X_n$ también son variables aleatorias cuyas distribuciones marginales son iguales a las de la población. Es decir, puede demostrarse que aun a pesar de que los objetos que se extraen de la población no sean reemplazados, la distribución no condicional de X_i es idéntica a la de la población, para toda $i = 1, 2, \ldots, n$.

La diferencia básica entre las dos técnicas es la noción de independencia. En la técnica dos, las observaciones $X_1, X_2, \ldots, X_n$ constituyen un conjunto de variables aleatorias independientes e idénticamente distribuidas (IID) dado que, por el proceso de reemplazo, ninguna observación se ve afectada por otra. En la técnica tres, a pesar de que las observaciones $X_1, X_2, \ldots, X_n$ poseen la misma distribución, no son independientes.

Recuérdese que, para la técnica uno, el muestreo se lleva a cabo con reemplazo a pesar de que la población no se encuentre constituida por objetos tangibles. De hecho, la técnica de muestreo dos es un caso especial de la primera, dado que la población no se afecta después de cada extracción. Sin embargo, es interesante notar que puede preferirse el muestreo aleatorio sin reemplazo si el tamaño de la población es relativamente pequeño*. En estos casos, si el muestreo se lleva a cabo con re-

* El lector recordará que esto es precisamente lo que constituye una distribución hipergeométrica tal como se discutió en la sección 4.4.

emplazo es muy probable que el mismo objeto sea seleccionado más de una vez. Es por esta razón que en las encuestas de preferencia el muestreo se hace sin reemplazo. Por otro lado, si el número de objetos en la población es muy grande, es irrelevante si el muestreo se lleva a cabo con reemplazo o sin éste. Conforme crece el tamaño de la población, el muestreo aleatorio sin reemplazo es, en todos los intentos y para cualquier propósito, igual al muestreo aleatorio con reemplazo.

Al hablar de la inferencia estadística se supondrá la existencia de una muestra aleatoria, como la descrita por la técnica de muestreo 1, y que se define de manera formal de la siguiente manera:

Definición 7.1 Si las variables aleatorias $X_1, X_2, \ldots, X_n$ tienen la misma función (densidad) de probabilidad que la de la distribución de la población y su función (distribución) conjunta de probabilidad es igual al producto de las marginales, entonces $X_1, X_2, \ldots, X_n$ forman un conjunto de n variables aleatorias independientes e idénticamente distribuidas (IID) que constituyen una *muestra aleatoria* de la población.

Cuando el objetivo es formular una inferencia estadística, debe hacerse un intento honesto para obtener una muestra aleatoria que porporcione la base teórica necesaria para la inferencia. Desde un punto de vista práctico, lo anterior no siempre es fácil. Por ejemplo, en muchas ocasiones es difícil decidir cuándo se están manteniendo condiciones idénticas durante el proceso de reunir datos en experimentos científicos. Esto es especialmente cierto si los factores ambientales crean condiciones heterogéneas. Sin embargo, es responsabilidad del experimentador decidir cuándo una muestra observada de datos es, en gran medida, aleatoria.

Para ilustrar el proceso de muestreo en un experimento científico, supóngase que se tiene interés en la concentración de cierto contaminante en un depósito de agua. Se coloca una boya que contiene un instrumento para medir el nivel de concentración en el sitio de interés. El instrumento registra el nivel de concentración cada n intervalos. De esta forma, las observaciones $X_1, X_2, \ldots, X_n$ constituyen una muestra del nivel de concentración en el sitio de interés. Antes de que el instrumento registre el nivel de concentración para el i-ésimo periodo, la observación X_i es una variable aleatoria para $i = 1, 2, \ldots, n$. El valor registrado x_i (el valor numérico correspondiente a la observación X_i) es una *realización* de la variable aleatoria. Al final de los n intervalos las mediciones $x_1, x_2, \ldots, x_n$ que registra el instrumento son las realizaciones, o datos muestrales, de las correspondientes variables aleatorias $X_1, X_2, \ldots, X_n$. Sin embargo, es válido preguntarse si la anterior es veraderamente una muestra aleatoria. Nadie puede proporcionar una respuesta legítima sin tener información adicional. Por ejemplo, ¿está el investigador consciente de todos los sucesos que durante el periodo de muestreo podría causar un cambio significativo en el nivel de concentración del contaminante? ¿Consideró el lapso de muestreo adecuado o existen algunas fluctuaciones temporales que deben ser consideradas? ¿Es probable que el error en el instrumento sea mayor conforme transcurre el tiempo? Preguntas como las anteriores deben contestarse antes de dar un juicio definitivo sobre la aleatoriedad de la muestra.

En el contexto de la definición 7.1, la función (densidad) conjunta de probabilidad de $X_1, X_2, \ldots, X_n$ es la función de verosimilitud de la muestra dada por

$$L(\underline{x}; \theta) = \prod_{i=1}^{n} f(x_i; \theta), \tag{7.1}$$

en donde $\underline{x} = \{x_1, x_2, \ldots, x_n\}$ denota los datos muestreados. Cuando las realizaciones $\underline{x}$ se conocen, $L(\underline{x}; \theta)$ es una función del parámetro desconocido θ. La utilidad de la función de verosimilitud para estimar parámetros se examinará en el capítulo ocho.

Ejemplo 7.1 Se ilustrará el concepto de muestra aleatoria dado en la definición 7.1 mediante lo siguiente: sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de n variables aleatorias IID de una población cuya distribución de probabilidad es exponencial con densidad

$$f(x; \theta) = \frac{1}{\theta}\exp(-x/\theta), \qquad 0 < x < \infty.$$

Cuando se observa X_1 y se registra su realización x_1,

$$f(x_1; \theta) = \frac{1}{\theta}\exp(-x_1/\theta), \qquad 0 < x_1 < \infty.$$

Ahora se observa X_2 y se registra su realización x_2. Dado que X_1 y X_2 son estadísticamente independientes y tienen las mismas densidades marginales,

$$f(x_2|x_1) = f(x_2; \theta) = \frac{1}{\theta}\exp(-x_2/\theta), \qquad 0 < x_2 < \infty.$$

La función de densidad conjunta de X_1 y X_2 es

$$f(x_1, x_2; \theta) = f(x_1; \theta) f(x_2; \theta) = \frac{1}{\theta^2}\exp[-(x_1 + x_2)/\theta], \qquad 0 < x_i < \infty, i = 1, 2.$$

Por lo tanto, se desprende que para una muestra aleatoria de tamaño n

$$L(x_1, x_2, \ldots, x_n; \theta) = \frac{1}{\theta^n}\exp[-(x_1 + x_2 + \cdots + x_n)/\theta],$$
$$0 < x_i < \infty, i = 1, 2, \ldots, n.$$

7.3 Distribuciones de muestreo de estadísticas

En los comentarios introductorios del capítulo uno se mencionó de manera breve que las características muestrales denominadas "estadísticas" se emplean para hacer inferencias con respecto a las características de la población, las que reciben el nombre de "parámetros". El objetivo de esta sección será el de examinar con detalle el papel que desempeñan las estadísticas en relación con la inferencia. En particular, se desa-

rrollará la noción de una distribución de muestreo de una estadística, que es uno de los conceptos más importantes en inferencia estadística.

Para colocar a las estadísticas en una mejor perspectiva se debe definir y analizar, de manera formal, un parámetro de población.

Definición 7.2 Un *parámetro* es una caracterización numérica de la distribución de la población de manera que describe, parcial o completamente, la función de densidad de probabilidad de la característica de interés. Por ejemplo, cuando se especifica el valor del parámetro de escala exponencial θ, se describe de manera completa la función de densidad de probabilidad

$$f(x; \theta) = \frac{1}{\theta} \exp(-x/\theta).$$

La oración "describe de manera completa" sugiere que una vez que se conoce el valor de θ entonces puede formularse cualquier proposición probabilística de interés. A manera de ilustración, si $\theta = 2$, entonces:

$$P(X > 4) = \frac{1}{2} \int_4^\infty \exp(-x/2)dx = 0.1353.$$

Por otra parte, si se especifica un valor del parámetro de forma α, de la distribución gama, la función de densidad de probabilidad

$$f(x; \alpha, \theta) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} \exp(-x/\theta)$$

no se encuentra especificada de manera completa, ya que no se ha hecho ninguna mención con respecto al valor del párametro de escala θ.

La esencia de todo lo anterior es que, dado que los parámetros son prácticamente inherentes a todos los modelos de probabilidad, es imposible calcular las probabilidades deseadas sin un conocimiento del valor de éstos. Es por esta razón que la noción de una estadística y su distribución de muestreo es muy importante en inferencia estadística. Esto es, los parámetros o sus funciones se estiman con base en estadísticas que, a su vez, se obtienen a partir de la información contenida en una muestra aleatoria.

Antes de dar la definición de una estadística, debe notarse que desde un punto de vista clásico (no bayesiano), un parámetro se considera como una constante fija cuyo valor se desconoce. Desde una perspectiva bayesiana un parámetro siempre es una variable aleatoria con algún tipo de distribución de probabilidad. Se considerará a los parámetros, principalmente desde el punto de vista clásico, aunque también se dará el punto de vista bayesiano, a fin de dar una perspectiva apropiada.

Definición 7.3 Una *estadística* es cualquier función de las variables aleatorias que se observaron en la muestra de manera que esta función no contiene cantidades desconocidas.

Considérese la muestra $\underline{X} = \{X_1, X_2, \ldots, X_n\}$ que consiste de n variables aleatorias IID con una función de densidad de probabilidad $f(x; \theta)$ que depende de un parámetro desconocido θ. Supóngase que se definen funciones como

$$T_1(\underline{X}) = (X_1 + X_2 + \cdots + X_n)/n,$$

$$T_2(\underline{X}) = (X_1^2 + X_2^2 + \cdots + X_n^2)/n,$$

$$T_3(\underline{X}) = X_1 + X_2,$$

y así sucesivamente. Todas ellas son estadísticas porque se determinan de manera completa por las variables aleatorias que contiene la muestra. De manera general, denótese una estadística por $T = u(\underline{X})$. Dado que T es una función de variables aleatorias, es en sí misma una variable aleatoria, y su valor específico $t = u(\underline{x})$ puede determinarse cuando se conozcan las realizaciones $\underline{x}$ de $\underline{X}$. Si se emplea una estadística T para estimar un parámetro desconocido θ, entonces T recibe el nombre de ***estimador*** de θ, y el valor específico de t como un resultado de los datos muestrales recibe el nombre ***estimación*** de θ. Esto es, un estimador es una estadística que identifica al mecanismo funcional por medio del cual, una vez que las observaciones en la muestra se realizan, se obtiene una estimación.

Una estadística es, sustancialmente, diferente de un parámetro. Un parámetro es una constante pero una estadística es una variable aleatoria. Además, un valor del parámetro descrito describe de manera completa un modelo de probabilidad (suponiendo una distribución uniparamétrica); ningún valor de la estadística puede desempeñar tal papel si cada uno de éstos depende del valor de las observaciones de las muestras. Y dado que las muestras se toman en forma aleatoria, ninguna muestra es más válida que cualquier otra que se haya tomado con el mismo fin.

Para ilustrar el concepto de una estadística se dará solución al siguiente problema: supóngase que se tiene interés en la duración promedio de cierta clase de batería miniatura. Se asegura que el proceso de manufactura de ésta es el mismo y que se emplean materiales idénticos. Se decide seleccionar aleatoriamente cinco pilas diarias durante 20 días. Para cada muestra diaria, las cinco baterías se someten a una prueba de duración que consiste en registrar el tiempo de operación. La prueba termina cuando todas dejan de funcionar. Como se supone que el proceso de fabricación es el mismo durante el periodo de muestreo, este esquema proporciona 20 muestras aleatorias distintas, donde cada una contiene cinco variables aleatorias independientes y distribuidas de manera idéntica. Sea $= \{X_{1j}, X_{2j}, \ldots, X_{5j}\}$ el conjunto de variables aleatorias de la $j$ésima muestra para $j = 1, 2, \ldots, 20$, y $\underline{x}_j = \{x_{1j}, x_{2j}, \ldots, x_{5j}\}$ los correspondientes tiempos de duración observados. Considérese la estadística.

$$T_j = (X_{1j} + X_{2j} + \cdots + X_{5j})/5$$

como un estimador del tiempo de duración promedio de las baterías. Si se supone que los tiempos observados son los que aparecen en la tabla 7.1, entonces para la j-ésima muestra existe una realización t_j para la estadística T_j. Es decir, cada muestra diaria proporciona una estimación de la duración promedio de las baterías.

Nótese que las estimaciones que aparecen en la tabla para la duración promedio tienen una variación que se encuentra entre 140.8 y 157.2 horas. De esta forma, existe una variabilidad inherente entre estas estimaciones. Además, para cualquier estadística se espera una variabilidad de muestra a muestra, dado que una estadística es una variable aleatoria. De hecho, para cada estadística existe lo que se conoce como su distribución de muestreo, la cual toma en cuenta la variabilidad inherente y proporciona los medios necesarios por medio de los cuales puede evaluarse la estadística. Se definirá la distribución de muestreo de una estadística con base en muestras aleatorias, de acuerdo con la definición 7.1.

Definición 7.4 La *distribución de muestreo* de una estadística T es la distribución de probabilidad de T que puede obtenerse como resultado de un número infinito de muestras aleatorias independientes, cada una de tamaño n, provenientes de la población de interés.

Dado que se supone que las muestras son aleatorias, la distribución de una estadística es un tipo de modelo de probabilidad conjunta para variables aleatorias independientes, en donde cada variable posee una función de densidad de probabilidad igual a la de las demás. De manera general, la distribución de muestreo de una estadística no tiene la misma forma que la función de densidad de probabilidad en la distribución de la población.

Para ilustrar lo anterior, considérese la distribución de muestreo de una estadística para los 20 promedios muestrales dados en la tabla 7.1. Mediante el empleo de los métodos del capítulo uno, se agrupan las 20 realizaciones en cinco clases y se obtienen las frecuencias relativas que aparecen en la tabla 7.2.

TABLA 7.1 Tiempos de duración (en horas) observados para una muestra aleatoria de baterías

Número de muestra	1	2	3	4	5	6	7	8	9	10
	163	159	150	136	136	138	155	158	135	166
	132	144	125	157	146	145	145	150	144	142
	154	139	139	168	158	150	151	153	148	156
	152	146	134	158	154	138	154	151	150	154
	148	144	156	167	156	158	141	138	148	160
Promedio de la muestra	149.8	146.4	140.8	157.2	150.0	145.8	149.2	150.0	145.0	155.6
Número de muestra	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
	150	154	148	149	150	147	158	164	153	135
	152	150	166	158	138	151	147	136	160	150
	163	141	148	139	153	161	141	143	156	164
	161	159	149	146	151	142	130	137	142	152
	139	153	154	136	161	149	147	152	156	144
Promedio de la muestra	153.0	151.4	153.0	145.6	150.6	150.0	144.6	146.4	153.4	149.0

TABLA 7.2 Grupos y frecuencias relativas para las 20 medias muestrales

Límites de clase	*Frecuencia de la clase*	*Frecuencia relativa*
140.6–144.0	1	0.05
144.1–147.5	6	0.30
147.6–151.0	7	0.35
151.1–154.5	4	0.20
154.6–158.0	2	0.10
Total	20	1.00

A partir de estas frecuencias relativas es evidente que la más alta concentración de tiempos de duración promedio se encuentra entre 147.6 y 151 horas, en donde los tiempos de duración promedio por debajo de 144 horas o por encima de 154.6 tienen una probabilidad muy pequeña. La distribución de muestreo de una estadística hace posible este tipo de análisis de probabilidad, esencial para valorar el riesgo inherente cuando se formulan inferencias.

Posteriormente se enunciarán algunos teoremas básicos que permiten obtener las distribuciones muestrales de estadísticas importantes como la media $\overline{X}$ y la varianza S^2 muestral. Se usará de manera frecuente la función generadora de momentos, dado que ésta determina unívocamente una distribución de probabilidad.

Teorema 7.1 Sea $X_1, X_2, \ldots, X_n$ un conjunto de n variables aleatorias independientes cada una con funciones generadoras de momentos $m_{X_1}(t)$, $m_{X_2}(t)$, ..., $m_{X_n}(t)$. Si

$$Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n,$$

en donde a_1, a_2, ..., a_n son constantes, entonces:

$$m_Y(t) = m_{X_1}(a_1t)m_{X_2}(a_2t) \cdots m_{X_n}(a_nt).$$

Demostración: Mediante el empleo de la definición y la hipótesis de independencia, se tiene

$$\begin{aligned} m_Y(t) &= E\{\exp[t(a_1X_1 + a_2X_2 + \cdots + a_nX_n)]\} \\ &= E[\exp(ta_1X_1)\exp(ta_2X_2) \cdots \exp(ta_nX_n)] \\ &= E[\exp(ta_1X_1)]E[\exp(ta_2X_2)] \cdots E[\exp(ta_nX_n)] \\ &= m_{X_1}(a_1t)m_{X_2}(a_2t) \cdots m_{X_n}(a_nt). \end{aligned}$$

De esta forma, la función generadora de momentos de una combinación lineal de n variables aleatorias independientes es el producto de las correspondientes funciones generadoras de momentos con argumentos iguales a las constantes de tiempo t.

Teorema 7.2 Sea $X_1, X_2, \ldots, X_n$ un conjunto de variables aleatorias independientes normalmente distribuidas con medias $E(X_i) = \mu_i$ y varianzas $Var(X_i) = \sigma_i^2$ para $i = 1, 2, \ldots, n$. Si

$$Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n,$$

en donde $a_1, a_2, \ldots, a_n$ son constantes, entonces Y es una variable aleatoria con distribución normal y media

$$E(Y) = a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n$$

y con varianza

$$Var(Y) = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2.$$

Demostración: Dado que X_i se encuentra normalmente distribuida, su función generadora de momentos es

$$m_{X_i}(t) = \exp[\mu_i t + (\sigma_i^2 t^2)/2].$$

De acuerdo con el teorema 7.1, la función generadora de momentos de Y es

$$\begin{aligned} m_Y(t) &= m_{X_1}(a_1t)m_{X_2}(a_2t) \cdots m_{X_n}(a_nt) \\ &= \exp[\mu_1 a_1 t + (a_1^2\sigma_1^2 t^2)/2] \cdots \exp[\mu_n a_n t + (a_n^2\sigma_n^2 t^2)/2] \\ &= \exp\left[t\sum_{i=1}^{n} a_i\mu_i + \left(t^2\sum_{i=1}^{n} a_i^2\sigma_i^2\right)\Big/2\right]. \end{aligned}$$

Por lo tanto, Y se encuentra normalmente distribuida con media $\Sigma_{i=1}^n a_i\mu_i$ y varianza $\Sigma_{i=1}^n a_i^2\sigma_i^2$.

Del teorema 7.2 se desprende que si $a_i = 1$ para $i = 1, 2, \ldots n$, entonces la suma de variables aleatorias independientes normalmente distribuidas también posee una distribución normal con media y varianza igual a la suma de las medias y las varianzas de cada una de las variables aleatorias. La mayor parte de las veces el resultado anterior se conoce como la *propiedad aditiva* de la distribución normal. Debe notarse que la hipótesis de normalidad no es necesaria para obtener las fórmulas de la media y la varianza de Y en el teorema 7.2. De hecho, con base en el teorema 6.1, si $X_1, X_2, \ldots, X_n$ es un conjunto de n variables aleatorias IID con medias $E(X_i) = \mu_i$ y varianzas $Var(X_i) = \sigma_i^2$, $i = 1, 2, \ldots, n$, entonces para $Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n$,

$$E(Y) = \sum_{i=1}^{n} a_i\mu_i$$

y (7.2)

$$Var(Y) = \sum_{i=1}^{n} a_i^2\sigma_i^2,$$

en donde, de nuevo, $a_1, a_2, \ldots, a_n$ son constantes.

Del teorema 7.2 surgen algunas aplicaciones interesantes. La siguiente constituye un ejemplo típico.

Ejemplo 7.2 Supóngase que para un árbol de levas y un cojinete, el diámetro externo del primero X_1 y el diámetro interno del segundo X_2 son variables aleatorias independientes con una distribución normal, con medias $E(X_1) = 3.25$ cm, $E(X_2) = 3.3$ cm y desvaciones estándar $d.e.(X_1) = 0.005$ cm y $d.e.(X_2) = 0.006$ cm, respectivamente. El interés recae en la diferencia entre X_2 y X_1, que es el espacio que existe entre el diámetro interno del cojinete y el diámetro externo del árbol de levas. El espacio se representa por Y, donde $Y = X_2 - X_1$. Si al armarse una máquina existe un apareamiento aleatorio entre los árboles de levas y los cojinetes, debe obtenerse el valor del espacio que existe entre éstos $y_{0.004}$, de manera tal que la probabilidad de que Y tenga un valor menor que éste sea de 0.004.

Dado que X_1 y X_2 son variables aleatorias independientes, se aplica el teorema 7.2 con $a_1 = -1$ y $a_2 = 1$. De esta forma

$$E(Y) = (1)E(X_2) + (-1)E(X_1) = 0.05,$$

y

$$d.e.(Y) = \sqrt{(1)^2(0.006)^2 + (-1)^2(0.005)^2} = 0.00781.$$

Esto es, $Y \sim N(0.05, 0.00781)$. Entonces

$$P(Y < y_{0.004}) = 0.004$$

o

$$P[Z < (y_{0.004} - 0.05)/0.00781] = 0.004,$$

pero

$$P[Z < -2.65] = 0.004;$$

así pues

$$(y_{0.004} - 0.05)/0.00781 = -2.65,$$

y $y_{0.004}$. De acuerdo con lo anterior se necesita un espacio no menor de 0.0293 cm para las condiciones dadas.

7.4 La distribución de muestreo de $\overline{X}$

Una de las estadísticas más importantes es la media de un conjunto de variables aleatorias independientes e idénticamente distribuidas. Esta estadística tiene un papel muy importente en problemas de toma de decisiones para medias poblacionales desconocidas. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria que consiste de n variables aleatorias IID tales que $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$ para toda $i = 1, 2, \ldots, n$. Entonces la estadística

$$\overline{X} = (X_1 + X_2 + \cdots + X_n)/n \tag{7.3}$$

se define como la media de las n variables aleatorias IID o, sencillamente, media muestral. Nótese que una vez que se conocen las realizaciones $x_1, x_2, ..., x_n$ de $X_1, X_2, ..., X_n$, respectivamente, la realización $\bar{x}$ de $\overline{X}$ se obtiene promediando los datos muestrales. Si en (7.2) $a_i = 1/n, i = 1, 2, ..., n$ entonces el valor esperado y la varianza de $\overline{X}$ son

$$E(\overline{X}) = \sum_{i=1}^{n} \frac{1}{n}\mu = n(\mu/n) = \mu \tag{7.4}$$

y

$$Var(\overline{X}) = \sum_{i=1}^{n} \frac{1}{n^2}\sigma^2 = n(\sigma^2/n^2) = \sigma^2/n, \tag{7.5}$$

respectivamente, en donde μ y σ^2 son la media y la varianza de la distribución de la población a partir de la cual se obtuvo la muestra. Con respecto a este resultado, lo importante es recordar que es válido sin importar la distribución de probabilidad de la población de interés siempre y cuando la varianza tenga un valor finito. A partir de (7.4), la desviación estándar de $\overline{X}$ es

$$d.e.\,(\overline{X}) = \sigma/\sqrt{n}\,, \tag{7.6}$$

la cual, en algunas ocasiones, recibe el nombre de *error estándar de la media.*

Nótese que conforme el tamaño de la muestra crece, la desviación estándar, y de esta forma la variabilidad, de $\overline{X}$ disminuye. En otras palabras, si el tamaño de la muestra crece, la precisión de la media muestral para estimar la media poblacional aumenta. Por ejemplo, si se extrae una muestra aleatoria de $n = 25$, $\overline{X}$ deberá tener una precisión de $\sqrt{25} = 5$ veces más de estimar la media poblacional que la que tendría una sola observación. Lo anterior es una propiedad muy ventajosa de la estadística $\overline{X}$ dado que asegura que para una muestra relativamente grande, se espera que la realización de $\overline{X}$ se encuentre muy cercana a la media poblacional μ. Como ilustración adicional, supóngase que se calcula la desviación estándar de $\overline{X}$ para distintos valores de n con $\sigma = 10$ y se grafican los puntos resultantes, como se indica en la figura 7.1. Por la naturaleza de 7.6, la desviación estándar de $\overline{X}$ sufre una disminución sustancial en su valor conforme n toma valores cada vez más grandes, pero si n es mayor de 30 o 40 este comportamiento cesa. Por lo tanto, en esencia, un tamaño grande de muestra no resulta razonable en cuanto al costo, si se hacen inferencias con respecto a μ con base en $\overline{X}$.

A continuación se enuncia y demuestra un teorema con respecto a la distribución de muestreo de $\overline{X}$ si la muestra se encuentra constituida por n variables aleatorias independientes normalmente distribuidas.

Teorema 7.3 Sea $X_1, X_2, ..., X_n$ una muestra aleatoria que consiste de n variables aleatorias independientes normalmente distribuidas con medias $E(X_i) = \mu$ y varianzas $Var(X_i) = \sigma^2$, $i = 1, 2, ..., n$. Entonces la distribución de la media muestral $\overline{X}$ es normal con media μ y varianza σ^2/n.

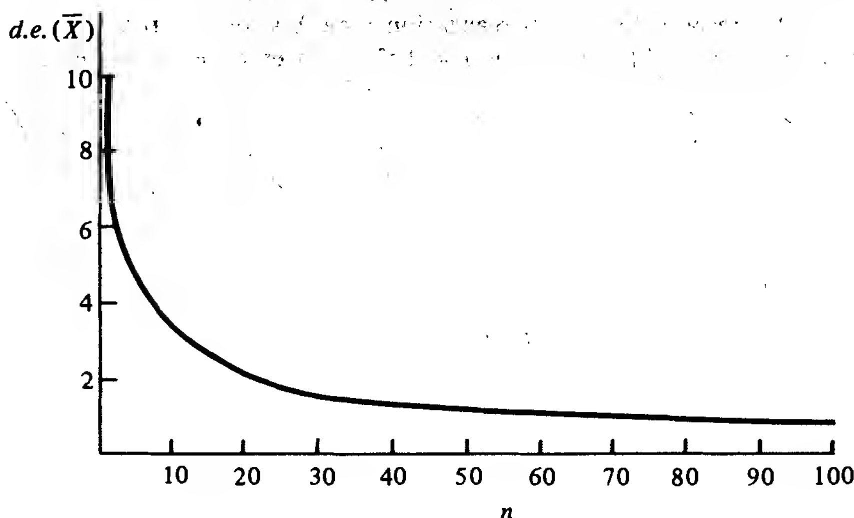


FIGURA 7.1 Comportamiento típico de la desviación estándar de $\overline{X}$ como función del tamaño de la muestra

Demostración: Este teorema es un corolario del teorema 7.2. Esto es, sea $a_i = 1/n$; dado que las medias y las varianzas son iguales, respectivamente, la función generadora de momentos de $\overline{X}$ es:

$$m\overline{X}(t) = \exp\left[t\sum_{i=1}^{n}\frac{1}{n}\mu + \left(t^2\sum_{i=1}^{n}\frac{1}{n^2}\sigma^2\right)\Big/2\right]$$

$$= \exp[\mu t + (t^2\sigma^2)/2n],$$

que es la función generadora de momentos de una variable aleatoria normalmente distribuida con media μ y varianza σ^2/n. De esta forma, la función de densidad de probabilidad de $\overline{X}$ cuando se muestrea una población cuya distribución es normal, está dada por

$$f(\bar{x}; \mu, \sigma/\sqrt{n}) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma}\exp\left[-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right], \qquad -\infty < \bar{x} < \infty. \qquad (7.7)$$

Ejemplo 7.3 Se tiene una máquina de llenado para vaciar 500 gr de cereal en una caja de cartón. Supóngase que la cantidad de cereal que se coloca en cada caja es una variable aleatoria normalmente distribuida con media 500 gr y desviación estándar igual a 20 gr. Para verificar que el peso promedio de cada caja se mantiene en 500 gr se toma una muestra aleatoria de 25 de éstas en forma periódica y se pesa el conteni-

do de cada caja. El gerente de la planta ha decidido detener el proceso y encontrar la falla cada vez que el valor promedio de la muestra sea mayor de 510 gr o menor de 490 gr. Obtener la probabilidad de detener el proceso.

Sean $X_1, X_2, \ldots, X_{25}$ variables aleatorias independientes normalmente distribuidas, las cuales representan la cantidad de cereal contenido en las cajas de una muestra aleatoria dada. Por hipótesis $X_i \sim N(500, 20)$, $i = 1, 2, \ldots, 25$. Por el teorema 7.3, el promedio muestral $\overline{X}$ también se encuentra normalmente distribuido con media 500 y desviación estándar $20/\sqrt{25} = 4$. La probabilidad deseada es igual a uno menos la probabilidad de que $\overline{X}$ se encuentre entre 490 y 510 gr; de esta forma

$$\begin{aligned} P(\text{Detención del proceso}) &= 1 - P(490 < \overline{X} < 510) \\ &= 1 - P\left(\frac{490 - 500}{4} < Z < \frac{510 - 500}{4}\right) \\ &= 1 - P(-2.5 < Z < 2.5) \\ &= 0.0124. \end{aligned}$$

Ejemplo 7.4 Demostrar que si $X_1, X_2, \ldots, X_n$ son n variables aleatorias independientes exponencialmente distribuidas con función de densidad de probabilidad

$$f(x; \theta) = \frac{1}{\theta}\exp(-x/\theta) \qquad x > 0,$$

entre $\overline{X}$ tiene una distribución gama.

Recuérdese que la función generadora de momentos de una variable aleatoria exponencialmente distribuida es $(1 - \theta t)^{-1}$. De esta forma, para cada X_i de la muestra,

$$m_{X_i}(t) = (1 - \theta t)^{-1}.$$

Del teorema 7.1 con $a_i = 1/n$, $i = 1, 2, \ldots, n$, se desprende que la función generadora de momentos de la media muestral $\overline{X}$ es

$$\begin{aligned} m_{\overline{X}}(t) &= m_{X_1}(t/n)m_{X_2}(t/n) \cdots m_{X_n}(t/n) \\ &= [1 - (\theta t/n)]^{-1}[1 - (\theta t/n)]^{-1} \cdots [1 - (\theta t/n)]^{-1} \\ &= [1 - (\theta t/n)]^{-n}. \end{aligned}$$

Pero la expresión anterior es la función generadora de momentos de una distribución gama con parámetro de forma n y parámetro de escala θ/n. De acuerdo con lo anterior, cuando se muestrea una población cuya distribución de probabilidad es exponencial, la densidad de probabilidad de $\overline{X}$ está dada por

$$f(\bar{x}; n, \theta/n) = \frac{n^n}{\Gamma(n)\theta^n}\bar{x}^{n-1}\exp(-n\bar{x}/\theta), \qquad \bar{x} > 0. \tag{7.8}$$

Nótese que si en las expresiones (5.47) y (5.48) se reemplaza α con n y θ con θ/n se obtiene

$$E(\overline{X}) = n\frac{\theta}{n} = \theta \tag{7.9}$$

y

$$Var(\overline{X}) = n\frac{\theta^2}{n^2} = \theta^2/n, \tag{7.10}$$

como era de esperarse ya que θ y θ^2 son la media y la varianza, respectivamente, de una variable aleatoria con distribución exponencial.

De la sección 5.5, recuérdese que si el parámetro de forma de una distribución gama tiene un valor grande, entonces los valores de probabilidad para una variable aleatoria gama pueden aproximarse, en forma adecuada, por una distribución normal. Dado que r^m, muestrear una distribución exponencial con parámetro θ $\overline{X}$ tiene una distribución gama con media θ, y desviación estándar $\theta/\sqrt{n}$, entonces, para n grande

$$Z = \frac{\overline{X} - \theta}{\theta/\sqrt{n}} \tag{7.11}$$

es, en forma aproximada, $N(0, 1)$.

Ejemplo 7.5 Con base en los experimentos, la duración de un componente eléctrico se encuentra exponencialmente distribuida con una vida media de 100 horas. Si del proceso de producción se toma una muestra aleatoria de 16 componentes, ¿cuál es la probabilidad de que la vida media muestral sea mayor de 120 horas?

De (7.9) y (7.10), la media de $\overline{X}$ en 100 horas y la desviación estándar tiene un valor de $100/\sqrt{16} = 25$ horas. Si se supone que el valor del parámetro de forma $n = 16$ es suficientemente grande para emplear la aproximación dada por (7.11), se tiene

$$P(\overline{X} > 120) = P\left(Z > \frac{120 - 100}{25}\right) = 0.2119.$$

Por comparación, la probabilidad de que $\overline{X} > 120$ pueda calcularse mediante el empleo directo de la función gama incompleta $I(u, p)$, se encuentra definida por (5.55); en este caso $u = (16)(120)/100\sqrt{16}$ y $p = 16 - 1$. De esta forma:

$$P(\overline{X} > 120) = 1 - I(4.8, 15) = 0.2021.$$

De manera muy breve se estableció ya que la distribución de muestreo de $\overline{X}$ es normal cuando éste se lleva a cabo a partir de una población que tiene una distribución, ya sea normal o exponencial. ¿Qué ocurre cuado no puede especificarse la distribución de probabilidad de la población a partir de la cual se obtiene la muestra? Es decir, ¿cuál es la distribución de muestreo (aproximada) de $\overline{X}$, sin tener en cuenta

la de las variables aleatorias de la muestra? Para obtener una idea con respecto a la distribución de muestreo de $\overline{X}$ cuando el modelo de probabilidad de la población de interés no se especifica, considérese un estudio de simulación en el que los valores aleatorios se generan mediante los procedimientos dados en la sección 5.9.

Supóngase que se generan 50 muestras, cada una de tamaño $n = 10$, a partir de una distribución de Poisson con parámetro $\lambda = 2$. Para cada muestra se calcula la media muestral, produciéndose así 50 realizaciones de la estadística $\overline{X}$. Estos valores se agrupan y se determinan sus frecuencias relativas. Se repite el proceso pero con $n = 40$ como tamaño de la muestra en lugar 10. Se repite el proceso pero en lugar de generar valores aleatorios a partir de una distribución de Poisson, se generan a partir de una distribución uniforme sobre el intervalo (0,1). En la figura 7.2 se ilustra la distribución de frecuencia relativa para cada uno de los cuatro casos. Nótese que cuando $n = 10$, no existe un patrón típico en la distribución de $\overline{X}$. Sin embargo, cuando $n = 40$ la distribución de $\overline{X}$ definitivamente toma una forma de campana y de esta forma se procede a una distribución normal, tanto para el modelo de Poisson como para el uniforme.

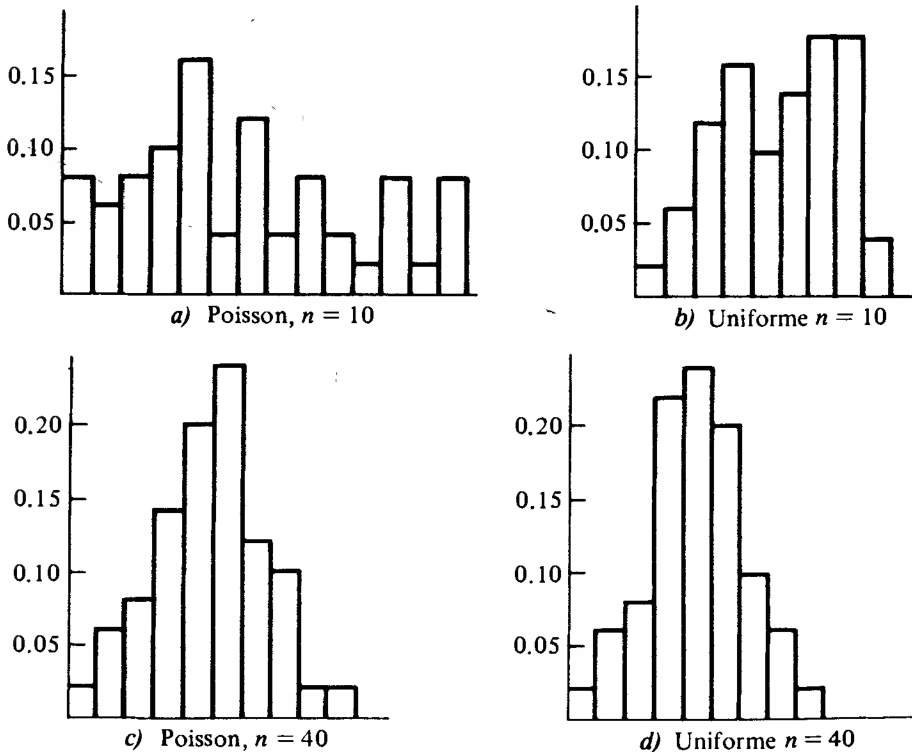


FIGURA 7.2 Distribuciones de frecuencia relativa de $\overline{X}$ cuando el muestreo se lleva a cabo sobre una distribución de Poisson o una uniforme para $n = 10$ y $n = 40$

Con base en este limitado estudio de simulación, parece ser que para un valor grande de n, la distribución de $\overline{X}$ es aproximadamente normal. De hecho, no importa el tipo de modelo de probabilidad a partir del cual se obtenga la muestra; mientras la media y la varianza existan, la distribución de muestreo de $\overline{X}$ se encontrará aproximada por $N(\mu, \sigma/\sqrt{n})$ para valores grandes de n.

Lo anterior constituye uno de los más importantes teoremas en inferencia estadística, y se conoce como *teorema central del límite.*

Teorema 7.4 Sean $X_1, X_2, \ldots, X_n$ n variables aleatorias IID con una distribución de probabilidad no especificada y que tienen una media μ y una varianza σ^2 finita. El promedio muestral $\overline{X} = (X_1 + X_2 + \cdots + X_n)/n$ tiene una distribución con media μ y varianza σ^2/n que tiende hacia una distribución normal conforme n tiende a ∞. En otras palabras, la variable aleatoria $(\overline{X} - \mu)/(\sigma/\sqrt{n})$ tiene como límite una distribución normal estándar. (En un apéndice al final de este capítulo se proporciona un esbozo de la demostración de este teorema.)

La esencia del teorema central del límite recae en el hecho de que para n grande, la distribución de $(\overline{X} - \mu)/(\sigma/\sqrt{n})$ es, en forma aproximada, normal con media cero y desviación estándar uno sin importar cuál sea el modelo de probabilidad a partir del que se obtuvo la muestra. Debe notarse que si el modelo de probabilidad de la población es semejante a una distribución normal (esto es, si es simétrico y existe una concentración relativamente alta alrededor del punto de simetría), la aproximación normal será buena aun para pequeñas muestras. Por otro lado, si el modelo de la población tiene muy poco parecido a una distribución normal (por ejemplo, existe una alta asimetría), la aproximación normal sólo sera adecuada para valores relativamente grandes de n. En muchos casos, puede concluirse de forma segura, que la aproximación será buena mientras $n > 30$. Por lo tanto, la variable aleatoria

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{7.12}$$

se emplea para formular inferencias acerca de μ cuando se conoce el valor de la varianza poblacional σ^2. La variable Z es $N(0, 1)$ cuando el muestreo se lleva a cabo sobre una población que tiene una distribución normal y es, en forma aproximada, $N(0, 1)$ para cualquier otro modelo cuando n es grande.

Ejemplo 7.6 Supóngase que el número de barriles de petróleo crudo que produce un pozo diariamente es una variable aleatoria con una distribución no especificada. Si se observa la producción en 64 días, seleccionados en forma aleatoria, y si se sabe que la desviación estándar del número de barriles por día es $\sigma = 16$, determínese la probabilidad de que la media muestral se encuentre a no más de cuatro barriles del verdadero valor de la producción por día.

Puesto que n es lo suficientemente grande, la distribución de $\overline{X}$ es, en forma aproximada, normal con media μ y desviación estándar $\sigma/\sqrt{n} = 16/\sqrt{64} = 2$. En

forma equivalente, la distribución de $Z = (\overline{X} - \mu)/2$ es, aproximadamente, $N(0, 1)$. De acuerdo con lo anterior, la probabilidad deseada es:

$$\begin{aligned} P(|\overline{X} - \mu| < 4) &= P(\mu - 4 < \overline{X} < \mu + 4) \\ &= P[(\mu - 4 - \mu)/2 < Z < (\mu + 4 - \mu)/2] \\ &= P(-2 < Z < 2) \\ &= 0.9544. \end{aligned}$$

7.5 La distribución de muestreo de S^2

Otra estadística importante empleada para formular inferencias con respecto a las varianzas de la población es la varianza muestral denotada por S^2. Recuérdese que S^2 es una medida de la variabilidad e indica la dispersión o extensión entre las observaciones. Dado que la dispersión es una consideración tan importante como la tendencia central, el significado de S^2 para formular inferencias de σ^2 es comparable con el que tiene $\overline{X}$ para formular inferencias con respecto a μ.

En esta sección se desarrollará la distribución de muestreo de S^2 cuando éste se lleva a cabo sobre una población que tiene una distribución normal. Para comenzar, es necesario suponer que μ es conocida y σ^2 no. Así, S^2 se encuentra definida por

$$S^2 = \sum_{i=1}^{n} (X_i - \mu)^2/n, \tag{7.13}$$

en donde $X_1, X_2, \ldots, X_n$ constituye una muestra aleatoria de una distribución normal con media μ y varianza σ^2 desconocida. Para determinar una distribución de muestreo que permita hacer inferencias sobre σ^2 con base en S^2 definida por (7.13), se enuncia y demuestra el siguiente teorema.

Teorema 7.5 Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución normal con media μ y varianza σ^2. La distribución de la variable aleatoria.

$$Y = \sum_{i=1}^{n} (X_i - \mu)^2/\sigma^2$$

es del tipo chi-cuadrada con n grados de libertad.

Demostración: Dado que $X_i \sim N(\mu, \sigma)$, $i = 1, 2, \ldots, n$, $Z_i = (X_i - \mu)/\sigma$ define n variables aleatorias normales estándar independientes, se tiene:

$$Y = \sum_{i=1}^{n} Z_i^2.$$

Del teorema 7.1,

$$m_Y(t) = m_{Z_1^2}(t)m_{Z_2^2}(t) \cdots m_{Z_n^2}(t)$$
$$= (1 - 2t)^{-1/2}(1 - 2t)^{-1/2} \cdots (1 - 2t)^{-1/2},$$

dado que el cuadrado de una variable aleatoria normal estándar tiene una distribución chi-cuadrada con un grado de libertad (véase el ejemplo 5.14). De esta forma se tiene

$$m_Y(t) = (1 - 2t)^{-n/2},$$

que es la función generadora de momentos de una distribución chi-cuadrada con n grados de libertad. De acuerdo con lo anterior, $Y \sim X_n^2$.

Ejemplo 7.7 Considérese una medición física proporcionada por un instrumento de precisión, en donde el interés recae en la variabilidad de la lectura. Supóngase que, con base en la experiencia, la medición es una variable aleatoria normalmente distribuida con media 10 y desviación estándar igual a 0.1 unidades. Si se toma una muestra aleatoria procedente del proceso de manufactura de los instrumentos de tamaño 25, ¿cuál es la probabilidad de que el valor de la varianza muestral sea mayor de 0.014 unidades cuadradas?

Con base en el teorema 7.5, la probabilidad de que $S^2 > 0.014$, cuando el muestreo se lleva a cabo sobre $N(10, 0.1)$ con $n = 25$ es igual a la de

$$P(Y > ns^2/\sigma^2) = P[Y > (25)(0.014)/0.01]$$
$$= P(Y > 35)$$
$$= 1 - P(Y \leq 35)$$

en donde $Y \sim X_{25}^2$. De la tabla E del apéndice, el valor de $P(Y \leq 35)$ es, aproximadamente, 0.9; de esta forma

$$P(Y > 35) \approx 0.1,$$

y la probabilidad de que el valor de la varianza muestral sea mayor de 0.014 unidades cuadradas, es alrededor de 0.1 para las condiciones dadas.

Desde un punto de vista práctico, la varianza muestra tal como se encuentra definida por (7.13) tiene poco uso, ya que es muy raro que se conozca el valor de la media poblacional μ. De acuerdo con lo anterior, si se muestra una distribución normal con media μ y varianza σ^2, la varianza muestral se define por

$$S^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2/(n - 1). \qquad (7.14)$$

En el capítulo ocho se verá por qué se emplea el divisor $(n - 1)$. El reemplazo de la media desconocida μ por la muestral $\overline{X}$ da origen a la presencia de otra estadística en la definición de S^2. De esta manera, para determinar la distribución de muestreo de

S^2, como se encuentra definida por (7.14), y con base en una muestra aleatoria proveniente de una distribución normal, debe tomarse en cuenta el promedio de la muestra $\overline{X}$. Como resultado se tiene que la distribución de muestreo de $(n-1)S^2/\sigma^2$ es también una distribución chi-cuadrada con $n - 1$ grados de libertad. A fin de probar lo anterior, primero se demostrará un teorema muy útil que involucra la suma de dos variables aleatorias independientes chi-cuadrada y entonces se escribe la expresión (7.14) en una forma equivalente, con objeto de aprovechar este teorema.

Teorema 7.6 Si X_1 y X_2 son dos variables aleatorias independientes y cada una tiene una distribución chi-cuadrada con ν_1 y ν_2 grados de libertad respectivamente, entonces:

$$Y = X_1 + X_2$$

también tiene una distribución chi-cuadrada con $\nu_1 + \nu_2$ grados de libertad.

Demostración: del teorema 7.1, la función generadora de momentos de Y es

$$\begin{aligned} m_Y(t) &= m_{X_1}(t)m_{X_2}(t) \\ &= (1-2t)^{-\nu_1/2}(1-2t)^{-\nu_2/2} \\ &= (1-2t)^{-(\nu_1+\nu_2)/2}, \end{aligned}$$

que es la función generadora de momentos de una variable aleatoria chi-cuadrada con $\nu_1 + \nu_2$ grados de libertad.

Ahora se deducirá la distribución de muestreo de $(n-1)S^2/\sigma^2$; de (7.14) se tiene que

$$(n-1)S^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2;$$

pero

$$\begin{aligned} \sum_{i=1}^{n} (X_i - \overline{X})^2 &= \sum_{i=1}^{n} (X_i - \mu - \overline{X} + \mu)^2 \\ &= \sum_{i=1}^{n} [(X_i - \mu) - (\overline{X} - \mu)]^2 \\ &= \sum_{i=1}^{n} [(X_i - \mu)^2 - 2(X_i - \mu)(\overline{X} - \mu) + (\overline{X} - \mu)^2] \\ &= \sum_{i=1}^{n} (X_i - \mu)^2 - 2(\overline{X} - \mu)\sum_{i=1}^{n} (X_i - \mu) + n(\overline{X} - \mu)^2 \\ &= \sum_{i=1}^{n} (X_i - \mu)^2 - 2(\overline{X} - \mu)n(\overline{X} - \mu) + n(\overline{X} - \mu)^2 \\ &= \sum_{i=1}^{n} (X_i - \mu)^2 - n(\overline{X} - \mu)^2. \end{aligned}$$

De esta forma

$$(n-1)S^2 + n(\overline{X}-\mu)^2 = \sum_{i=1}^{n} (X_i - \mu)^2.$$

Al dividir ambos miembros de la expresión anterior por la varianza poblacional σ^2 se tiene

$$\frac{(n-1)S^2}{\sigma^2} + \frac{n(\overline{X}-\mu)^2}{\sigma^2} = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\sigma^2},$$

o

$$\frac{(n-1)S^2}{\sigma^2} + \left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2 = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\sigma^2}. \qquad (7.15)$$

Del teorema 7.15, se desprende que $\sum_{i=1}^{n} (X_i - \mu)^2/\sigma^2$ tiene una distribución chi-cuadrada con n grados de libertad. De manera similar, $[(\overline{X} - \mu)/\sigma/\sqrt{n}]^2$ también posee una distribución chi-cuadrada con un grado de libertad, dado que $(\overline{X} - \mu)/(\sigma/\sqrt{n})$ es $N(0, 1)$. Por lo tanto, si se supone que $(n-1)S^2/\sigma^2$ y $[(\overline{X} - \mu)/\sigma/\sqrt{n}]^2$son variables aleatorias independientes, entonces, por el teorema 7.6, cuando se muestrea una población cuya distribución es normal con media y varianza desconocida, la distribución de $(n-1)S^2/\sigma^2$, es chi-cuadrada con n – grados de libertad. Para demostrar la independencia se invita al lector a que consulte la referencia [3]. La función de densidad de probabilidad de $Y = (n-1)S^2/\sigma^2$ se desprende de (5.58) y está dada por:

$$f(y; n-1) = \begin{cases} \dfrac{1}{\Gamma[(n-1)/2]2^{(n-1)/2}} y^{[(n-1)/2]-1} \exp(-y/2) & y > 0, \\ 0 & \text{para cualquier otro valor.} \end{cases} \qquad (7.16)$$

Nótese que, dado que $Y \sim X^2_{n-1}$, $E(Y) = n - 1$ y $Var(Y) = 2(n-1)$. Además, ya que $Y = (n-1)S^2/\sigma^2$, $S^2 = \sigma^2 Y/(n-1)$. Por lo tanto

$$E(S^2) = E[\sigma^2 Y/(n-1)] = \frac{\sigma^2}{(n-1)} E(Y) = \sigma^2, \qquad (7.17)$$

y

$$Var(S^2) = Var[\sigma^2 Y/(n-1)] = \frac{\sigma^4}{(n-1)^2} Var(Y) = \frac{2\sigma^4}{n-1}. \qquad (7.18)$$

7.6 La distribución t de Student

Se recordará de la sección 7.5 que cuando se muestrea una distribución normal con desviación estándar conocida σ, la distribución de $Z = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ es $N(0,$

1). Desde un punto de vista práctico, la necesidad de conocer σ impide formular inferencias con respecto a μ debido a que generalmente no se conoce el valor de la desviación estándar de la población. Dada la disponibilidad de una muestra aleatoria, el camino lógico que se sigue en este caso es reemplazar σ con una estimación s, que es el valor de la desviación estándar muestral S. Desafortunadamente, cuando lo anterior se lleva a cabo, la distribución de $(\bar{X} - \mu)/(S/\sqrt{n})$ no es $N(0, 1)$, aun cuando la muestra provenga de una distribución normal. Sin embargo, es posible determinar la distribución de muestreo exacta de $(\bar{X} - \mu)/(S/\sqrt{n})$ cuando se muestrea $N(\mu, \sigma)$, con μ y σ^2 desconocidos. Para finalizar esta sección se examinarán los aspectos teóricos de lo que se conoce como la distribución t de Student.*

Supóngase que se realiza un experimento en que se observan dos variables aleatorias X y Z; X tiene una distribución chi-cuadrada con ν grados de libertad y Z una distribución normal con media cero y desviación estándar uno. Sea T otra variable aleatoria que es función de X y Z, de manera tal que

$$T = \frac{Z}{\sqrt{X/\nu}}. \tag{7.19}$$

Es decir, T se define como el cociente entre una variable aleatoria normal estándar y la raíz cuadrada de una variable aleatoria chi-cuadrada dividida por sus grados de libertad. El conjunto de todos los posibles valores de la variable aleatoria T es el intervalo $(-\infty, \infty)$ puesto que los valores de Z se encuentran en éste y los valores de X son positivos. El valor

$$t = \frac{z}{\sqrt{x/\nu}}$$

recibe el nombre de valor de la variable aleatoria de t de Student. Lo anterior lleva al siguiente teorema.

Teorema 7.7 Sea Z una variable aleatoria normal estándar y X una variable aleatoria chi-cuadrada con ν grados de libertad. Si Z y X son independientes, entonces la variable aleatoria

$$T = \frac{Z}{\sqrt{X/\nu}}$$

tiene una distribución t de Student con ν grados de libertad y una función de densidad de probabilidad dada por

$$f(t; \nu) = \frac{\Gamma[(\nu + 1)/2]}{\sqrt{\pi\nu}\,\Gamma(\nu/2)}[1 + (t^2/\nu)]^{-(\nu+1)/2}, \qquad -\infty < t < \infty, \quad \nu > 0. \tag{7.20}$$

La deducción de la función de densidad t de Student aparece en un apéndice al final de este capítulo.

De (7.20) se observa que el parámetro de la distribución t es ν, que, al igual que para la distribución chi-cuadrada, recibe el nombre de grados de libertad. Para cual-

* W. Gosset, desarrolló en 1908 la distribución t, quien publicó su trabajo bajo el seudónimo de "Student".

quier $\nu > 0$, la distribución t es simétrica con respecto al origen y la función de densidad tiene su valor máximo cuando $t = 0$. De la figura 7.3 es evidente que la forma de la función de densidad t de Student es muy similar a la de la densidad normal estándar y con los extremos de la distribución t menos pronunciados que los de la distribución normal. De hecho, conforme se tiene un número mayor de grados de libertad, la distribución t de Student tiende hacia la normal estándar.

Puede demostrarse que el valor esperado de T es

$$E(T) = 0 \qquad \nu > 1, \tag{7.21}$$

y la varianza está dada por

$$Var(T) = \nu/(\nu - 2) \qquad \nu > 2. \tag{7.22}$$

En la tabla F del apéndice se encuentran los valores cuantiles $t_{1-\alpha,\nu}$ tales que:

$$P(T \leq t_{1-\alpha,\nu}) = \int_{-\infty}^{t_{1-\alpha,\nu}} f(t; \nu)dt = 1 - \alpha, \qquad 0 \leq \alpha \leq 1, \tag{7.23}$$

para los distintos valores de ν y de las proporciones acumulativas seleccionadas $1 - \alpha$. Por ejemplo, si $\nu = 15$.

$$P(T \leq t_{0.90,15}) = P(T \leq 1.341) = 0.90,$$

$$P(T \leq t_{0.95,15}) = P(T \leq 1.753) = 0.95,$$

$$P(T \leq t_{0.99,15}) = P(T \leq 2.602) = 0.99.$$

Dado que la distribución t es simétrica con respecto al cero, para $\alpha > 0.5$ los valores cuantiles $t_{1-\alpha,\nu}$ serán negativos pero sus magnitudes serán las mismas que las

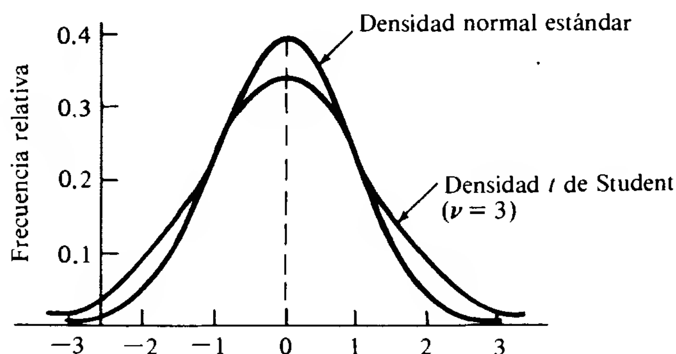


FIGURA 7.3 Comparación entre las densidades normal estándar y t de Student

de los correspondientes valores que se encuentran en el lado derecho. De esta forma, para $\nu = 15$,

$$P(T \leq t_{0.10,15}) = P(T \leq -1.341) = 0.10,$$

$$P(T \leq t_{0.05,15}) = P(T \leq -1.753) = 0.05,$$

$$P(T \leq t_{0.01,15}) = P(T \leq -2.602) = 0.01.$$

A fin de ilustrar la similitud que existe entre la distribución t de Student y la normal estándar para valores relativamente grandes de ν, en la tabla 7.3 se encuentra una comparación entre los valores cuantiles t y los correspondientes valores normales estándar para valores crecientes de ν. Para $\alpha = 0.1$ o 0.05, la concordancia se encuentra en aproximadamente 0.05 unidades, aun para valores tan bajos de ν como 30. De hecho, muchos autores sugieren que, desde un punto de vista práctico, es muy poca la ganancia que se tiene al emplear la distribución t de Student en lugar de la normal estándar cuando $\nu \geq 30$.

Recuérdese que para formular inferencias con respecto a μ cuando el muestreo se lleva a cabo sobre una distribución normal con media y varianza desconocidas, se necesita determinar la distribución de $(\overline{X} - \mu)/(S/\sqrt{n})$. Cuando se muestrea una distribución $N(\mu, \sigma)$ se sabe, del teorema 7.3, que la distribución de $(\overline{X} - \mu)/(\sigma/\sqrt{n})$ es $N(0, 1)$. Para la misma condición, se sabe que, de (7.15) y del teorema 7.6, la distribución de $(n - 1)S^2/\sigma^2$ es chi-cuadrada con $n - 1$ grados de libertad. Dado que puede demostrarse que $\overline{X}$ y S^2 son independientes, del teorema 7.7 se desprende que la distribución de

$$\frac{\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n - 1)S^2/\sigma^2}{(n - 1)}}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \cdot \frac{\sigma}{\sqrt{S^2}},$$

o

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}, \tag{7.24}$$

es la t de Student con $n - 1$ grados de libertad.

TABLA 7.3 Comparación entre los valores cuantiles de las distribuciones t de Student y normal estándar

α	$t_{1-\alpha,20}$	$t_{1-\alpha,30}$	$t_{1-\alpha,40}$	$t_{1-\alpha,50}$	$z_{1-\alpha}$
0.10	1.325	1.310	1.303	1.299	1.282
0.05	1.725	1.697	1.684	1.676	1.645
0.01	2.528	2.457	2.423	2.403	2.326

Ejemplo 7.8 El Departamento de Protección al Medio Ambiente asegura que, para un automóvil compacto en particular, el consumo de gasolina en carretera es de un galón por cada 45 millas. Una organización independiente de consumidores adquiere uno de estos sutomóviles y lo somete a prueba con el propósito de verificar la cifra proporcionada por el DPMA. El automóvil recorrió una distancia de 100 millas en 25 ocasiones. En cada recorrido se anotó el número de galones necesarios para realizar el viaje. Los 25 ensayos, el valor promedio y la desviación estándar, tuvieron un valor de 43.5 y 2.5 millas por galón, respectivamente. Si se supone que el número de millas que se recorre por galón es una variable aleatoria distribuida normalmente, con base en esta prueba ¿existe alguna razón para dudar de la varacidad del dato proporcionado por el DONA?

Este problema ilustra algunas de las dificultades prácticas que pueden encontrarse al ponerse en práctica la noción de muestra aleatoria. En forma ideal, se debieron seleccionar 25 carros de la misma marca, modelo y configuración de motor, de manera aleatoria, del mismo proceso de armado, de manera que fuese posible considerar el consumo de combustible como una variable aleatoria. Sin embargo, en éste y otros, lo anterior representa un costo prohibitivo. A pesar de lo anterior, debe determinarse la veracidad de la información proporcionada por el DPMA con base en la probabilidad. Esto es, si μ fuese realmente igual a 45 millas por galón, ¿Cuál es la probabilidad de que se observe un valor de $\bar{X}$ no mayor de 43.5 millas por galón, con base en una muestra de tamaño 25 y una estimación de σ igual a 2.5?

De (7.24) puede verse que

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{43.5 - 45}{2.5/\sqrt{25}}$$

$$= -3$$

es un valor de la distribución t de Student con 24 grados de libertad. De la tabla F del apéndice se tiene que $P(T \leq -3) < 0.005$. Es decir, si el valor verdadero de la media es 45, la probabilidad de observar un valor de T no mayor de -3 unidades, es menor de 0.005. En cualquier caso, se ha observado algo que tiene una posibilidad de ocurrir menos de 5 en 1 000, o μ tiene un valor real menor de 45. Para esta situación es preferible elegir la segunda explicación.

7.7 La distribución de la diferencia entre dos medias muestrales

En muchas ocasiones surge la necesidad de comparar las medias de dos distribuciones distintas. Por ejemplo, supóngase que se tiene interés en comparar los tiempos de duración promedio de las baterías para automóvil "48 meses" de las marcas Mears and Sawbuck y J.C. Nickel. Las baterías vendidas por dos comerciantes, de manera factible, se producen por compañías distintas y se fabrican bajo diferentes especificaciones. Para cada una se supondrá que existe una distribución, diferente de la otra, que toma en cuenta la duración de las baterías.

Sea X una variable aleatoria que representa la duración del acumulador Mears and Sawbuck, en forma que $X \sim N(\mu_X, \sigma)$. De manera similar, sea Y la correspondiente variable aleatoria para las baterías J.C. Nickel tal que $Y \sim N(\mu_Y, \sigma)$. Nótese que se supone que las varianzas de X y Y son iguales. Se selecciona una muestra aleatoria de n_X baterías de la marca Mears and Sawbuck y una muestra aleatoria de n_Y de la marca J.C. Nickel. Los acumuladores de las dos muestras se someten a la misma prueba de duración en la que se controlan todos los factores externos identificados. Las diferencias observadas para los tiempos de duración en ambas marcas se deben sólo a la variabilidad inherente del proceso de fabricación respectivo. El interés recae en formular una inferencia con respecto a la diferencia $\mu_X - \mu_Y$ entre las dos medias desconocidas.

Un enfoque viable para este problema es formular la inferencia con base en la diferencia que hay entre las dos medias muestrales $\bar{X}$ y $\bar{Y}$. De acuerdo con lo anterior, se necesita obtener la distribución de $\bar{X} - \bar{Y}$ cuando el muestreo se lleva a cabo sobre dos poblaciones normales independientes con varianzas iguales. Si se supone que el valor de la varianza σ^2 se conoce del teorema 7.3, se sabe que la distribución de $\bar{X}$ es normal con media μ_X y varianza σ^2/n_X. La distribución de $\bar{Y}$ también es normal pero con media μ_Y y varianza σ^2/n_Y. Dado que $\bar{X}$ y $\bar{Y}$ son variables aleatorias independientes normalmente distribuidas, si $a_1 = 1$ y $a_2 = -1$ en el teorema 7.2, la distribución de $\bar{X} - \bar{Y}$ también es normal con media $\mu_X - \mu_Y$ y varianza $(\sigma^2/n_X) + (\sigma^2/n_Y) = \sigma^2(1/n_X + 1/n_Y)$. Por lo tanto, si se conoce el valor de σ^2, la distribución de

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sigma\sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}} \tag{7.25}$$

es $N(0, 1)$. La expresión (7.25) proporciona un camino adecuado por medio del cual se puede formular una inferencia con respecto a la diferencia de las medias poblacionales de dos distribuciones normales independientes con igual varianza.

En el desarrollo de (7.25) se supuso que el valor de σ^2 era conocido. Sin embargo, es poco probable conocer el valor de σ^2 para una situación real. Así pues, debe obtenerse la distribución de $\bar{X} - \bar{Y}$ cuando el muestreo se lleve a cabo sobre dos poblaciones normales independientes con varianzas iguales pero desconocidas. Para cada una de las dos muestras aleatorias, pueden definirse las varianzas muestrales S_X^2 y S_Y^2 dadas por (7.14). Dado que $(n_X - 1)S_X^2/\sigma^2$ y $(n_Y - 1)S_Y^2/\sigma^2$ son dos variables independientes chi-cuadrada, con $n_X - 1$ y $n_Y - 1$ grados de libertad respectivamente, por el teorema 7.6, la distribución de

$$W = \frac{(n_X - 1)S_X^2}{\sigma^2} + \frac{(n_Y - 1)S_Y^2}{\sigma^2} \tag{7.26}$$

también es chi-cuadrada con $n_X + n_Y - 2$ grados de libertad. De la expresión (7.19) se desprende el hecho de que el cociente de Z en (7.25) y la raíz cuadrada de W divida entre sus grados de libertad tiene una distribución t de Student con $n_X + n_Y - 2$

grados de libertad. Esto es,

$$\frac{[\bar{X} - \bar{Y} - (\mu_X - \mu_Y)]/\sigma\sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}}{\sqrt{\dfrac{[(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2]/\sigma^2}{n_X + n_Y - 2}}} = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\dfrac{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}{n_X + n_Y - 2}\left(\dfrac{1}{n_X} + \dfrac{1}{n_Y}\right)}},$$

o

$$T = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_p\sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}}, \tag{7.27}$$

en donde

$$S_p^2 = [(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2]/(n_X + n_Y - 2) \tag{7.28}$$

que, en general, recibe el nombre de estimador combinado (pooled) de la varianza común σ^2. Nótese de (7.28) que S_p^2 es el promedio, con factores de peso, de las dos varianzas muestrales S_X^2 y S_Y^2, siendo los factores de peso los grados de libertad. De acuerdo con lo anterior, se puede formular una inferencia con respecto a la diferencia entre μ_X y μ_Y con base en (7.27), cuando el muestreo se lleva a cabo sobre dos poblaciones cuyas distribuciones son anormales e independientes y en donde las varianzas son iguales pero sus valores no se conocen.

En este momento es natural que el lector pregunte qué pasa si no es posible suponer que la varianza de las dos distribuciones sea la misma. Si las varianzas σ_X^2 y σ_Y^2 no son iguales, pero se conocen sus valores, el problema es sencillo. La distribución de

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}} \tag{7.29}$$

aún es $N(0, 1)$, por las mismas razones que llevaron a la expresión (7.25). Por otro lado, si se desconocen los valores de las varianzas y además éstos no son iguales, el problema es mucho más complicado y por esta razón no debe emplearse la expresión (7.27). En esencia, una situación como la anterior constituye lo que se conoce como el problema de Fisher-Behrens, el cual se encuentra más allá del alcance de este libro. Existen algunas aproximaciones a este problema, una de la cuales puede encontrarse en [1].

7.8 La distribución *F*

De la sección 7.5, recuérdese que las inferencias con respecto a σ^2 cuando se muestrea una distribución normal, se formulan con base en la estadística $(n - 1)S^2/\sigma^2$, la que tiene una distribución chi-cuadrada con $n - 1$ grados de libertad. En esta sección se desarrollará la estadística apropiada para emplearse en la formulación de

inferencias con respecto a las varianzas de dos distribuciones normales independientes con base en las muestras aleatorias de cada una. Por último, se analizará la teoría de una distribución muy útil, la cual se conoce como distribución F.

Supóngase un experimento en que se observan dos variables aleatorias independientes X y Y, cada una con una distribución chi-cuadrada con ν_1 y ν_2 grados de libertad respectivamente. Sea F una variable aleatoria que es función de X y Y, de manera tal que

$$F = \frac{X/\nu_1}{Y/\nu_2}. \qquad (7.30)$$

Esto es, la variable aleatoria F es el cociente de dos variables aleatorias chi-cuadrada, cada una dividida por sus grados de libertad. Lo anterior lleva al siguiente teorema.

Teorema 7.8 Sean X y Y dos variables aleatorias independientes chi-cuadrada con ν_1 y ν_2 grados de libertad, respectivamente. La variable aleatoria

$$F = \frac{X/\nu_1}{Y/\nu_2}$$

tiene una distribución F con una función de densidad de probabilidad dada por

$$g(f; \nu_1, \nu_2)^* = \begin{cases} \dfrac{\Gamma[(\nu_1 + \nu_2)/2]\nu_1^{\nu_1/2}\nu_2^{\nu_2/2}}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} f^{(\nu_1-2)/2}(\nu_2 + \nu_1 f)^{-(\nu_1+\nu_2)/2} & f > 0, \\ 0 & \text{para cualquier otro valor} \end{cases} \qquad (7.31)$$

(La deducción de la función de densidad de probabilidad de F es similar a la de la t de Student y se deja como ejercicio para el lector.)

La distribución F se caracteriza completamente por los grados de libertad ν_1 y ν_2. Puede demostrarse que el valor esperado es

$$E(F) = \nu_2/(\nu_2 - 2) \qquad \nu_2 > 2, \qquad (7.32)$$

y la varianza está dada por

$$Var(F) = \frac{\nu_2^2(2\nu_2 + 2\nu_1 - 4)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)} \qquad \nu_2 > 4. \qquad (7.33)$$

La distribución F tiene asimetría positiva para cualesquiera valores de ν_1 y ν_2, pero ésta va disminuyendo conforme ν_1 y ν_2 toman valores cada vez más grandes.

En la tabla G del apéndice, se encuentran los valores cuantiles $f_{1-\alpha,\nu_1,\nu_2}$, tales que

$$P(F \leq f_{1-\alpha,\nu_1,\nu_2}) = \int_0^{f_{1-\alpha,\nu_1,\nu_2}} g(f; \nu_1, \nu_2)df = 1 - \alpha, \qquad 0 \leq \alpha \leq 1 \qquad (7.34)$$

* Se emplea g para denotar la función de densidad y de esta forma evitar cualquier confusión con respecto al argumento f.

para las proporciones acumulativas seleccionadas $1 - \alpha$ y distintas combinaciones de los grados de libertad del numerador ν_1, y del denominador ν_2 del cociente (7.30). Por ejemplo, si $\nu_1 = 5$ y $\nu_2 = 10$, entonces:

$$P(F \leq f_{0.90,5,10}) = P(F \leq 2.52) = 0.90,$$

$$P(F \leq f_{0.95,5,10}) = P(F \leq 3.33) = 0.95,$$

$$P(F \leq f_{0.99,5,10}) = P(F \leq 5.64) = 0.99.$$

Nótese que en la tabla G se encuentran los valores cuantiles $f_{1-\alpha,\nu_1,\nu_2}$ únicamente para $\alpha < 0.5$. Si se desean los cuantiles del lado izquierdo (es decir, para $\alpha > 0.5$) éstos pueden encontrarse mediante el siguiente procedimiento: si la variable aleatoria F tiene una distribución F con ν_1 y ν_2 grados de libertad, entonces la variable $F' = 1/F$ también tiene una distribución F pero con ν_2 y ν_1 grados de libertad. Puede verse que lo anterior es cierto, a partir de (7.30),

$$F' = \frac{1}{\dfrac{X/\nu_1}{Y/\nu_2}} = \frac{Y/\nu_2}{X/\nu_1}. \tag{7.35}$$

Si se desean los valores cuantiles $f_{1-\alpha,\nu_1,\nu_2}$ para $\alpha > 0.5$,

$$P(F \leq f_{1-\alpha,\nu_1,\nu_2}) = P\left(\frac{1}{F} > \frac{1}{f_{1-\alpha,\nu_1,\nu_2}}\right) = 1 - \alpha,$$

o

$$P\left(\frac{1}{F} \leq \frac{1}{f_{1-\alpha,\nu_1,\nu_2}}\right) = \alpha. \tag{7.36}$$

Pero $1/F = F' \sim F$ se encuentra distribuida con ν_2 y ν_1 grados de libertad. Entonces el α-ésimo valor cuantil de F' es tal que

$$P(F' \leq f'_{\alpha,\nu_2,\nu_1}) = \alpha. \tag{7.37}$$

Dado que (7.36) y (7.37) son idénticas, se sigue que

$$f'_{\alpha,\nu_2,\nu_1} = 1/f_{1-\alpha,\nu_1,\nu_2}$$

y

$$f_{1-\alpha,\nu_1,\nu_2} = 1/f'_{\alpha,\nu_2,\nu_1} \qquad \text{for } \alpha > 0.5. \tag{7.38}$$

Como ejemplo, sea $\nu_1 = 8$ y $\nu_2 = 12$. Entonces

$$P(F \leq f_{0.05,8,12}) = P(F \leq 1/f'_{0.95,12,8}) = P(F \leq 1/3.28) = P(F \leq 0.305) = 0.05,$$

o

$$P(F \leq f_{0.01,8,12}) = P(F \leq 1/f'_{0.99,12,8}) = P(F \leq 1/5.67) = P(F \leq 0.176) = 0.01.$$

Regresando al problema de desarrollar una estadística apropiada para usarse en la formulación de inferencias con respecto a las varianzas de dos distribuciones normales independientes, sea $X_1, X_2, \ldots, X_{n_X}$ una muestra aleatoria de variables aleatorias independientes y normalmente distribuidas cada una con media μ_X y varianza σ_X^2. También sea $Y_1, Y_2, \ldots, Y_{n_Y}$ un conjunto de n_Y variables aleatorias independientes normalmente distribuidas, cada una con media μ_Y y varianza σ_Y^2. Si se supone que las X y las Y son independientes, las estadísticas

$$(n_X - 1)S_X^2/\sigma_X^2$$

y

$$(n_Y - 1)S_Y^2/\sigma_Y^2$$

son dos variables aleatorias chi-cuadrada independientes con $n_X - 1$ y $n_Y - 1$ grados de libertad, respectivamente. Entonces, por el teorema 7.8, se desprende que la variable aleatoria

$$\frac{\dfrac{(n_X - 1)S_X^2}{\sigma_X^2} \Big/ (n_X - 1)}{\dfrac{(n_Y - 1)S_Y^2}{\sigma_Y^2} \Big/ (n_Y - 1)} = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \tag{7.39}$$

tiene una distribución F con $n_X - 1$ y $n_Y - 1$ grados de libertad.

Una aplicación de (7.39) es inmediata si se recuerda el problema general de la sección 7.7. Esto es, el formular una inferencia con respecto a la diferencia entre dos medias poblacionales ya sea cuando se conocen las varianzas de las poblaciones o cuando se supone que se conoce, al menos, el cociente de éstas. Una forma factible de verificar la validez de esta suposición es mediante el empleo de (7.39). Si la suposición de que $\sigma_X^2 = \sigma_Y^2$ es correcta, la estadística F dada por (7.39), se reduce a

$$F = S_X^2/S_Y^2. \tag{7.40}$$

Cuando se obtienen los valores de S_X^2 y S_Y^2 a partir de las muestras y se calcula el cociente (7.40), puede concluirse que la hipótesis de varianza iguales es falsa si el valor de este cociente es, de manera suficiente, distinto de 1. En otras palabras, si las dos varianzas son iguales, la probabilidad de observar un valor de F distinto, de manera suficiente, es pequeña.

Para finalizar, debe notarse que en esta sección, así como en las secciones 7.5 y 7.7, se desarrolló el material que se presentó bajo la hipótesis de realizar un muestreo aleatorio sobre poblaciones que tienen una distribución normal. En la realidad, la hipótesis de normalidad puede o no ser justificable. Sin embargo, desde un punto de vista práctico, el lector debe darse cuenta que la diferencia entre la distribución normal y el modelo de probabilidad de la población de interés es inversamente proporcional a las técnicas delineadas para formular inferencias. La afirmación anterior es particularmente cierta cuando se formulan inferencias con respecto a las varianzas cuando se emplean la distribución chi-cuadrada o la F.

Referencias

1. P. G. Hoel, *Introduction to mathematical statistics*, 4th ed., Wiley, New York, 1971.
2. R. V. Hogg and A. T. Craig, *Introduction to mathematical statistics*, 4th ed., MacMillan, New York, 1978.
3. B. W. Lindgren, *Statistical theory*, 3rd ed., MacMillan, New York, 1976.
4. A. M. Mood and F. A. Graybill, *Introduction to the theory of statistics*, 2nd ed., McGraw-Hill, New York, 1963.

Ejercicios

7.1. Una firma de mercadotecnia envía un cuestionario a 1 000 residentes de cierto suburbio de una ciudad para determinar sus preferencias como compradores. De los 1 000 residentes, 80 responden el cuestionario. ¿Lo anterior constituye una muestra aleatoria? Discutir los méritos de este procedimiento para obtener una muestra aleatoria.

7.2. En una planta de armado automotriz se seleccionarán 50 de los primeros 1 000 automóviles de un nuevo modelo para ser inspeccionados por el departamento de control de calidad. El gerente de la planta decide inspeccionar un automóvil cada vez que terminan de armarse 20. ¿Este proceso dará como resultado una muestra aleatoria? Comente.

7.3. Si $X_1, X_2, \ldots, X_n$ constituye una muestra aleatoria, obtener las funciones de verosimilitud de las siguientes distribuciones:

a) De Poisson, con parámetro λ;
b) Hipergeométrica, con parámetro p;
c) Uniforme en el intervalo (a, b);
d) $N(\mu, \sigma)$.

7.4. Repetir el ejercicio 7.3 para las siguientes distribuciones:

a) Gama con parámetro α y θ,
b) Weibull con parámetro α y θ.

7.5. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población cuya distribución es normal con media μ y varianza σ^2 desconocidas. De las siguientes, ¿cuáles son estadísticas?

a) $\Sigma X_i - \mu$
b) $\sigma X_1 + \sigma X_2$
c) $X_i, i = 1, 2, \ldots, n$
d) $X_1^2 + X_2^2 - \exp(X_3)$
e) $X_i/\sigma, i = 1, 2, \ldots, n$
f) $\Sigma(X_i - \bar{X})^2$

7.6. Sean $X_1, X_2, \ldots, X_n$ n variables aleatorias independientes de Poisson con parámetros $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectivamente. Mediante el empleo de la función generadora de momentos, demostrar que la suma de estas variables también es una variable aleatoria de Poisson con parámetros $\lambda_1 + \lambda_2 + \cdots + \lambda_n$.

7.7. Sean X_1 y X_2 dos variables aleatorias independientes de Poisson con parámetros λ_1 y λ_2, respectivamente. Demostrar que la diferencia entre X_1 y X_2 no es una variable aleatoria de Poisson.

7.8. Sean X_1 y X_2 dos variables aleatorias independientes binomial con parámetros n_1 y p, y n_2 y p, respectivamente. Demostrar que la suma de X_1 y X_2 es una variable aleatoria binomial con parámetros $n_1 + n_2$ y p.

7.9. Sean X_1 y X_2 dos variables aleatorias independientes distribuidas exponencialmente con el mismo parámetro θ. Demostrar que la suma de X_1 y X_2 es una variable aleatoria gama con parámetro de forma 2 y parámetro de escala θ.

7.10. Para un determinado nivel de ingresos, el Departamento de Hacienda sabe que las cantidades declaradas por concepto de deducciones médicas (X_1), contribuciones caritativas (X_2) y gastos varios (X_3), son variables aleatorias independientes normalmente distribuidas con medias \$400, \$800 y \$100 y desviaciones estándar \$100, \$250 y \$40, respectivamente.

a) ¿Cuál es la probabilidad de que la cantidad total declarada por concepto de estas tres deducciones, no sea mayor de \$1 600?

b) Si una persona con este nivel de ingresos declara por concepto de estas deducciones un total de \$2 100, ¿qué tan probable es tener una cantidad igual o mayor a este monto bajo las condiciones dadas?

7.11. Una tienda de artículos eléctricos para el hogar vende tres diferentes marcas de refrigeradores. Sean X_1, X_2 y X_3 variables aleatorias las cuales representan el volumen de ventas mensual para cada una de las tres marcas de refrigeradores. Si X_1, X_2 y X_3 son variables aleatorias independientes normalmente distribuidas con medias \$8 000, \$15 000 y \$12 000, y desviaciones estándar \$2 000, \$5 000 y \$3 000, respectivamente, obtener la probabilidad de que, para un mes en particular, el volumen de venta total para los tres refrigeradores sea mayor de \$50 000.

7.12 En una tienda de servicio el tiempo total del sistema consta de dos componentes (el lapso de tiempo que debe esperarse para que el servicio dé comienzo (X_1) y el lapso de tiempo que éste dura (X_2)). Si X_1 y X_2 son variables aleatorias independientes exponencialmente distribuidas con un tiempo medio de 4 minutos cada una, ¿cuál es la probabilidad de que el tiempo total que tarda el sistema en proporcionar el servicio no sea mayor de 15 minutos? (Sugerencia: consulte el ejercicio 7.9.)

7.13. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población que tiene una distribución gama con parámetros α y θ. Mediante el uso de la función generadora de momentos, demostrar que la distribución de la media muestral $\overline{X}$ también es de tipo gama, con parámetros de escala y de forma iguales a $n\alpha$ y θ/n respectivamente.

7.14. Mediante el empleo de los resultados de la sección 5.9, generar números aleatorios para las distribuciones binomial y exponencial y usarlos para demostrar el teorema central del límite. De manera específica, para $n = 10$ y $n = 40$, generar 50 muestras de una distribución binomial con $p = 0.4$. Repetir el procedimiento anterior generando 50 muestras de una distribución exponencial con parámetro $\theta = 100$. ¿Se ha demostrado el teorema central del límite en un grado razonable?

7.15. Para cierta prueba de aptitud se sabe con base en la experiencia que el número de aciertos es 1 000 con una desviación estándar de 125. Si se aplica la prueba a 100 personas seleccionadas al azar, aproximar las siguientes probabilidades que involucran a la media muestral $\overline{X}$.

a) $P(985 < \overline{X} < 1015)$ *c)* $P(\overline{X} > 1020)$

b) $P(960 < \overline{X} < 1040)$ *d)* $P(\overline{X} < 975)$

7.16. Un contratista piensa comprar una gran cantidad de lámparas de alta intensidad a cierto fabricante. Éste asegura al contratista que la duración promedio de las lámparas es de

1 000 horas con una desviación estándar igual a 80 horas. El contratista decide comprar las lámparas sólo si una muestra aleatoria de 64 de éstas da como resultado una vida promedio de por lo menos 1 000 horas. ¿Cuál es la probabilidad de que el contratista adquiera las lámparas?

7.17. Un inspector federal de pesos y medidas visita una planta de empacado para verificar que el peso neto de las cajas sea el indicado en éstas. El gerente de la planta asegura al inspector que el peso promedio de cada caja es de 750 gr con una desviación estándar de 5 gr. El inspector selecciona, al azar, 100 cajas y encuentra que el peso promedio es de 748 gr. Bajo estas condiciones, ¿qué tan probable es tener un peso de 748 o menos? ¿Qué actitud debe tomar el inspector?

7.18. En la fabricación de cojinetes para motores, se sabe que el diámetro promedio es de 5 cm con una desviación estándar igual a 0.005 cm. El proceso es vigilado en forma periódica mediante la selección aleatoria de 64 cojinetes, midiendo sus correspondientes diámetros. El proceso no se detiene mientras la probabilidad de que la media muestral se encuentre entre dos límites especificados sea de 0.95. Determinar el valor de estos límites.

7.19. En la producción de cierto material para soldar se sabe que la desviación estándar de la tensión de ruptura de este material es de 25 libras. ¿Cuál debe ser la tensión de ruptura promedio del proceso si, con base en una muestra aleatoria de 50 especímenes, la probabilidad de que la media muestral tenga un valor mayor de 250 libras es de 0.95?

7.20. Genere 50 muestras, cada una de tamaño 25 a partir de una distribución normal con media 60 y desviación estándar 10. Calcule la varianza de cada muestra mediante el empleo de (7.14).

a) Obtener la media y la varianza de S^2 mediante el empleo de los 50 valores calculados. ¿Cómo son estos valores al compararlos con los proporcionados por las expresiones (7.17) y (7.18)?

b) Agrupar los 50 valores calculados de S^2 y graficar las frecuencias relativas. Coméntese sobre los resultados.

7.21. Repetir el ejercicio 7.20 pero generando los valores aleatorios a partir de una distribución exponencial con parámetro de escala $\theta = 30$. Haga un comentario sobre sus resultados.

7.22. Para un gerente de planta es muy importante controlar la variación en el espesor de un material plástico. Se sabe que la distribución del espesor del material es normal con una desviación estándar de 0.01 cm. Una muestra aleatoria de 25 piezas de este material da como resultado una desviación estándar muestral de 0.015 cm. Si la varianza de la población es $(0.01)^2$ cm^2, ¿cuál es la probabilidad de que la varianza muestral sea igual o mayor que $(0.015)^2$ cm^2? Por lo tanto, ¿qué puede usted concluir con respecto a la variación de este proceso?

7.23. Si se obtiene una muestra aleatoria de $n = 16$ de una distribución normal con media y varianza desconocidas, obtener $P(S^2/\sigma^2 \leq 2.041)$.

7.24. Si se obtiene una muestra aleatoria de tamaño $n = 21$ de una distribución normal con media y varianza desconocidas, obtener $P(S^2/\sigma^2 \leq 1.421)$.

7.25. Un fabricante de cigarrillos asegura que el contenido promedio de nicotina, en una de sus marcas, es de 0.6 mg por cigarrillo. Una organización independiente mide el contenido de nicotina de 16 cigarrillos de esta marca y encuentra que el promedio y la desvia-

ción estándar muestral es de 0.75 y 0.175 mg, respectivamente, de nicotina. Si se supone que la cantidad de nicotina en estos cigarrillos es una variable aleatoria normal, ¿qué tan probable es el resultado muestral dado el dato proporcionado por el fabricante?

7.26. Durante los 12 meses pasados el volumen diario de ventas de un restaurante fue de $2 000. El gerente piensa que los próximos 25 días serán típicos con respecto al volumen de ventas normal. Al finalizar los 25 días, el volumen de ventas y su desviación estándar promedio fueron de $1 800 y $200, respectivamente. Supóngase que el volumen de ventas diario es una variables aleatoria normal. Si usted fuese el gerente, ¿tendría alguna razón para creer, con base en este resultado, que hubo una disminución en el volumen de ventas promedio diario?

7.27. El gerente de una refinería piensa modificar el proceso para producir gasolina a partir de petróleo crudo. El gerente hará la modificación sólo si la gasolina promedio que se obtiene por este nuevo proceso (expresada como un porcentaje del crudo) aumenta su valor con respecto al proceso en uso. Con base en un experimento de laboratorio y mediante el empleo de dos muestras aleatorias de tamaño 12, una para cada proceso, la cantidad de gasolina promedio del proceso en uso es de 24.6 con una desviación estándar de 2.3, y para el proceso propuesto fue de 28.2 con una desviación estándar de 2.7. El gerente piensa que los resultados proporcionados por los dos procesos son variables aleatorias independientes normalmente distribuidas con varianzas iguales. Con base en esta evidencia, ¿debe adoptarse el nuevo proceso?

7.28. Una organización independiente está interesada en probar la distancia de frenado a una velocidad de 50 mph para dos marcas distintas de automóviles. Para la primera marca se seleccionaron nueve automóviles y se probaron en un medio controlado. La media muestral y la desviación estándar fueron de 145 pies y 8 pies, respectivamente. Para la segunda marca se seleccionaron 12 automóviles y la distancia promedio resultó ser de 132 pies y una desviación estándar de 10 pies. Con base en esta evidencia, ¿existe alguna razón para creer que la distancia de frenado para ambas marcas, es la misma? Supóngase que las distancias de frenado son variables aleatorias independientes normalmente distribuidas con varianzas iguales.

7.29. La variación en el número de unidades diarias de cierto producto, el cual manejan dos operadores A y B, debe ser la misma. Con base en muestras de tamaño $n_A = 16$ días y $n_B = 21$ días, el valor calculado de las desviaciones estándar muestrales es de $s_A = 8.2$ unidades y $s_B = 5.8$ unidades. Si el número de éstas, manejadas por los dos operadores, por día, son dos variables aleatorias independientes que se encuentran aproximadas, en forma adecuada, por distribuciones normales, ¿existe alguna razón para creer que las varianzas son iguales?

7.30. Con base en la información proporcionada en el ejercicio 7.27, ¿existe alguna razón para creer que las varianzas de los dos procesos son iguales?

APÉNDICE

Demostración del teorema central del límite

El propósito de este apéndice no es el presentar una demostración general y elegante desde el punto de vista matemático, sino más bien proporcionar un esbozo de la de-

mostración del teorema central del límite. Se quiere demostrar que la función generadora de momentos de $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ tiende a la de una distribución normal estándar conforme n tiende al infinito. Sean

$$Z_i = (X_i - \mu)/\sigma \qquad i = 1, 2, ..., n,$$

y

$$Y = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

Dado que

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma/\sqrt{n}}\right) = \frac{1}{n}\cdot\frac{1}{\sigma/\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu) = \frac{1}{n}\cdot\frac{1}{\sigma/\sqrt{n}}(n\bar{X} - n\mu) = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}},$$

entonces

$$Y = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i.$$

Como resultado se tiene que la función generadora de momentos de Y es igual a la función generadora de momentos de $(1/\sqrt{n})\,\Sigma_{i=1}^{n} Z_i$. Del teorema 7.1,

$$\begin{aligned} m_Y(t) &= [m_{Z_i}(t/\sqrt{n})]^n \\ &= \{E[\exp(tZ_i/\sqrt{n})]\}^n, \end{aligned}$$

dado que las Z_i son variables aleatorias independientes.

Al expander $(tZ_i/\sqrt{n})$ en una serie de Taylor:

$$\exp(tZ_i/\sqrt{n}) = 1 + \frac{t}{\sqrt{n}}Z_i + \frac{t^2}{2n}Z_i^2 + \frac{t^3}{3!n^{3/2}}Z_i^3 + \cdots.$$

Si se toman los valores esperados y se recuerda que $E(Z_i) = 0$ y $Var(Z_i) = 1$, $i = 1, 2, ..., n$, se tiene

$$E[\exp(tZ_i/\sqrt{n})] = 1 + \frac{t^2}{2n} + \frac{t^3}{3!n^{3/2}}E(Z_i^3) + \cdots.$$

De acuerdo con lo anterior

$$\begin{aligned} m_Y(t) &= \left[1 + \frac{t^2}{2n} + \frac{t^3}{3!n^{3/2}}E(Z_i^3) + \cdots\right]^n \\ &= \left\{1 + \frac{1}{n}\left[\frac{t^2}{2} + \frac{t^3}{3!\sqrt{n}}E(Z_i^3) + \cdots\right]\right\}^n \\ &= \left(1 + \frac{u}{n}\right)^n \end{aligned}$$

en donde

$$u = \frac{t^2}{2} + \frac{t^3}{3!\sqrt{n}}E(Z_i^3) + \cdots .$$

Ahora

$$\lim_{n\to\infty} m_Y(t) = \lim_{n\to\infty}\left(1 + \frac{u}{n}\right)^n,$$

pero por definición

$$\lim_{n\to\infty}\left(1 + \frac{u}{n}\right)^n = e^u.$$

Lo anterior da como resultado una situación idéntica a la que se tiene en la demostración del teorema 5.1. Esto es, conforme $n\to\infty$, todos los términos en u, excepto el primero, tienden hacia cero debido a que todos tienen potencias positivas de n en sus denominadores. Por lo tanto, puede deducirse que

$$\lim_{n\to\infty} m_Y(t) = \exp(t^2/2),$$

o la distribución límite de $Y = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ es la normal estándar para valores grandes de n.

APÉNDICE

Deducción de la función de densidad de probabilidad t de Student

Sea T una variable aleatoria definida por (7.19). Considere la densidad de probabilidad de T cuando X se mantiene fija en un valor x. Dado que

$$f_Z(z) = \frac{1}{\sqrt{2\pi}}\exp(-z^2/2),$$

la densidad de probabilidad condicional de

$$T = Z/(x/\nu)^{1/2}$$

se obtiene al considerar la relación inversa

$$Z = (x/\nu)^{1/2}T$$

y al sustituir en $f_Z(z)$, en donde el jacobiano de la transformación es

$$\frac{dz}{dt} = (x/\nu)^{1/2}.$$

De esta forma

$$f(t|x) = \frac{(x/\nu)^{1/2}}{\sqrt{2\pi}}\exp(-xt^2/2\nu), \qquad -\infty < t < \infty, \quad x > 0.$$

De (6.19) se sabe que la densidad conjunta de T y X es

$$f(t, x) = f(t|x)f_X(x).$$

Dado que $X \sim \chi^2_\nu$,

$$f_X(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)}x^{(\nu-2)/2}\exp(-x/2), \qquad x > 0.$$

De esta forma

$$\begin{aligned} f(t, x) &= \frac{1}{\sqrt{2\pi\nu}\, 2^{\nu/2}\Gamma(\nu/2)}x^{(\nu-1)/2}\exp\left(-\frac{x}{2} - \frac{xt^2}{2\nu}\right) \\ &= c_1 x^{(\nu-1)/2}\exp(-c_2 x/2), \end{aligned}$$

en donde $c_1 = 1/[\sqrt{2\pi\nu}\, 2^{\nu/2}\Gamma(\nu/2)]$ y $c_2 = [1 + (t^2/\nu)]$. Integrando $f(t, x)$ con respecto a x, se obtiene la función de densidad de probabilidad de la distribución t de Student. De acuerdo con lo anterior

$$\begin{aligned} f_T(t) &= c_1 \int_0^\infty x^{(\nu-1)/2}\exp(-c_2 x/2)dx \\ &= c_1 \int_0^\infty (2y/c_2)^{(\nu-1)/2}\exp(-y)(2/c_2)dy, \text{ en donde } y = c_2 x/2 \text{ y } dx = (2/c_2)dy \\ &= c_1(2/c_2)^{(\nu+1)/2} \int_0^\infty y^{(\nu-1)/2}\exp(-y)dy \\ &= c_1(2/c_2)^{(\nu+1)/2}\Gamma[(\nu + 1)/2] \\ &= \frac{1}{\sqrt{2\pi\nu}\, 2^{\nu/2}\Gamma(\nu/2)} \cdot \frac{2^{(\nu+1)/2}}{[1 + (t^2/\nu)]^{(\nu+1)/2}}\Gamma[(\nu + 1)/2] \\ &= \frac{\Gamma[(\nu + 1)/2]}{\sqrt{\pi\nu}\, \Gamma(\nu/2)}\left[1 + (t^2/\nu)\right]^{-(\nu+1)/2}, \qquad -\infty < t < \infty. \end{aligned}$$

CAPÍTULO OCHO

Estimación puntual y por intervalo

8.1 Introducción

En el capítulo anterior se mencionó, en forma breve, que las estadísticas se emplean para estimar los valores de parámetros desconocidos o funciones de éstos. En este capítulo se examinará con detalle el concepto de *estimación de parámetros* mediante la especificación de las propiedades deseables de los estimadores (estadísticas) y el desarrollo de técnicas apropiadas para implementar el proceso de estimación. Se utilizará el punto de vista de la teoría del muestreo, que considera a un parámetro como una cantidad fija pero desconocida.

La estimación de un parámetro involucra el uso de los datos muestrales en conjunción con alguna estadística. Existen dos formas de llevar a cabo lo anterior: la *estimación puntual* y la *estimación por intervalo.* En la primera se busca un estimador que, con base en los datos muestrales, dé origen a una estimación univaluada del valor del parámetro y que recibe el nombre de *estimado puntual.* Para la segunda, se determina un intervalo en el que, en forma probable, se encuentra el valor del parámetro. Este intervalo recibe el nombre de *intervalo de confianza estimado.*

Al igual que en los capítulos anteriores, la función de densidad de probabilidad en la distribución de la población de interés se denotará por $f(x;\ \theta)$, donde la función depende de un parámetro arbitrario θ, el cual puede tomar cualquier valor que se encuentre en cierto dominio.* De esta forma, el principal objetivo de este capítulo es presentar los criterios convenientes para la determinación de los estimadores de θ.

8.2 Propiedades deseables de los estimadores puntuales

Con el propósito de mostrar la necesidad de estimar parámetros, considérese la siguiente situación. Cuando se obtiene una muestra aleatoria de cierta característica X

* El dominio de un parámetro recibe el nombre de espacio parametral.

de la distribución de la población, y a pesar de que pueda identificarse la forma funcional de la densidad de ésta, es poco probable que la característica pueda especificarse de manera completa mediante los valores de todos los parámetros. En esencia, se conoce la familia de distribuciones a partir de la cual se obtiene la muestra, pero no puede identificarse el miembro de interés de ésta, ya que no se conoce el valor del parámetro. Este último tiene que estimarse con base en los datos de la muestra. Por ejemplo, supóngase que la distribución del tiempo de servicio en una tienda es exponencial con parámetro desconocido θ. Se observan 25 lapsos aleatorios y la media muestral calculada es igual a 3.5 minutos. Dado que para la distribucíñ exponencial $E(X) = \theta$, un estimado puntual de θ es 3.5. Por lo tanto, de manera aparente, el muestreo se llevó a cabo sobre una distribución exponencial cuya media estimada es de 3.5 minutos.

Es posible definir muchas estadísticas para estimar un parámetro desconocido θ. Por ejemplo, para el caso anterior pudo elegirse la mediana muestral para estimar el valor de la media. Entonces, ¿cómo seleccionar un buen estimador de θ? ¿Cuáles son los criterios para juzgar cuándo un estimador de θ es "bueno" o "malo"? De manera intuitiva, ¿qué es un buen estimador? Si se piensa en términos de "estimadores humanos" como los que se encuentran en las compañías grandes de construcción, entonces quizá un buen estimador sea aquella persona cuyas estimaciones siempre se encuentran muy cercanas a la realidad. Como ejemplo adicional, suponga que un grupo de personas se encuentra al tanto del volumen de ventas y adquisiciones de tres comerciantes (A, B y C) quienes compiten en el mismo mercado. Como el inventario es siempre un aspecto importante en los negocios, cada uno de estos comerciantes predice la demanda mensual de sus productos y, con base en ésta, realizan las adquisiciones necesarias. Supóngase que se determina la diferencia entre las demandas real y la esperada para varios meses y con base en éstas se obtienen las distribuciones de frecuencia que se muestran en la figura 8.1.

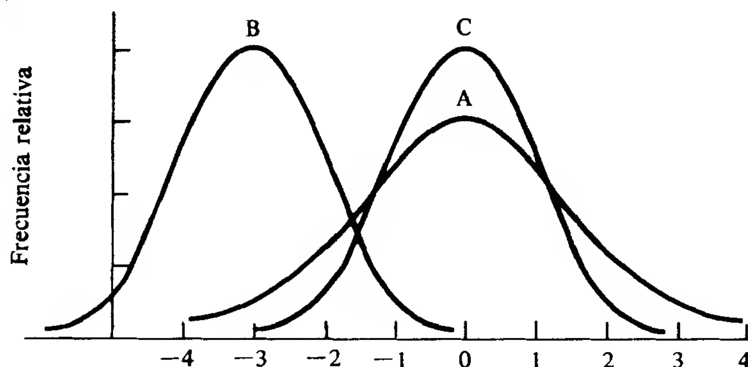


FIGURA 8.1 Frecuencias alisadas para la diferencia entre las demandas real y predecida

La intuición sugiere que el comerciante C es el que hace mejor su trabajo no sólo porque la distribución de la diferencia entre las demandas real y esperada se concentra alrededor del valor perfecto de cero sino también porque la variabilidad de la diferencia es, en forma relativa, pequeña. Para el comerciante A, aun a pesar de que la distribución también se encuentra centrada alrededor del origen, existe una mayor variabilidad en las diferencias. La distribución para el comerciante B se concentra alrededor de un valor negativo, lo cual sugiere que B sobreestima la mayor parte del tiempo la demanda mensual.

Si se acepta la premisa de que el objetivo de la estimación de parámetros no es igual al de los estimadores o predictores humanos, entonces, de los ejemplos anteriores, surgen dos propiedades deseables: el estimador de un parámetro θ debe tener una distribución de muestreo concentrada alrededor de θ y la varianza del estimador debe ser la menor posible.

Para ampliar las propiedades anteriores, considérese lo siguiente. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño n proveniente de una distribución con función de densidad $f(x; \theta)$, y sea $T = u(X_1, X_2, \ldots, X_n)$ cualquier estadística. El problema es encontrar una función u que sea la que proporcione la "mejor" estimación de θ. Al buscar el mejor estimador de θ se hará uso de una cantidad muy importante que recibe el nombre de error cuadrático medio de un estimador.

Definición 8.1 Sea T cualquier estimador de un parámetro desconocido θ. Se define el error cuadrático medio de T como el valor esperado del cuadrado de la diferencia entre T y θ.

Para cualquier estadística T, se denotará el error cuadrático medio por ECM(T); de esta forma

$$\text{ECM}(T) = E(T - \theta)^2. \tag{8.1}$$

Puede verse la razón del por qué el error cuadrático medio es una cantidad importante para enjuiciar a los posibles estimadores de θ mediante el desarrollo de (8.1); este es,

$$\begin{aligned} \text{ECM}(T) &= E(T^2 - 2\theta T + \theta^2) \\ &= E(T^2) - 2\theta E(T) + \theta^2 \\ &= Var(T) + [E(T)]^2 - 2\theta E(T) + \theta^2 \\ &= Var(T) + [\theta - E(T)]^2. \end{aligned} \tag{8.2}$$

El error cuadrático medio de cualquier estimador es la suma de dos cantidades no negativas: una es la varianza del estimador y la otra es el cuadrado del sesgo del estimador. El lector encontrará que estas dos cantidades se encuentran relacionadas en forma directa con las propiedades deseables de un estimador. De manera específica, la varianza de un estimador debe ser lo más pequeña posible mientras que la distribución de muestreo debe concentrarse alrededor del valor del parámetro. Por lo tanto, el problema visto de manera superficial parece bastante sencillo; esto es, seleccionar, como el mejor estimador de θ, la estadística que tenga el error cuadrático medio

más pequeño posible de entre todos los estimadores factibles de θ. Sin embargo, en realidad el problema es mucho más complicado. Aun si fuese práctico determinar los errores cuadráticos medios de un número grande de estimadores, para la mayor parte de las densidades $f(x;\ \theta)$ no existe ningún estimador que minimice el error cuadrático medio para todos los posibles valores de θ. Es decir, un estimador puede tener un error cuadrático medio mínimo para algunos valores de θ, mientras que otro estimador tendrá la misma propiedad, pero para otros valores de θ.

Ejemplo 8.1 **Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de alguna distribución tal que $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$, $i = 1, 2, \ldots, n$. Considere las estadísticas**

$$T_1 = \overline{X}$$

y

$$T_2 = \sum_{i=1}^{n} X_i/(n + 1)$$

como posibles estimadores de μ. Obtener los errores cuadráticos medios de T_1 y T_2 y demostrar que $\text{ECM}(T_2) < \text{ECM}(T_1)$ para algunos valores de μ mientras que la proposición inversa es cierta para otros valores de μ.

El sesgo de T_1 es cero, dado que $E(T_1) = E(\overline{X}) = \mu$; de esta forma se tiene

$$\text{ECM}(T_1) = Var(T_1) = \sigma^2/n.$$

Para T_2,

$$\begin{aligned} E(T_2) &= (n + 1)^{-1}E\left(\sum_{i=1}^{n} X_i\right) \\ &= (n + 1)^{-1}\sum_{i=1}^{n} E(X_i) \\ &= n\mu/(n + 1). \end{aligned}$$

De manera similar,

$$\begin{aligned} Var(T_2) &= Var\left[(n + 1)^{-1}\sum_{i=1}^{n} X_i\right] \\ &= (n + 1)^{-2}\sum_{i=1}^{n} Var(X_i) \\ &= n\sigma^2/(n + 1)^2. \end{aligned}$$

De esta forma se tiene

$$\begin{aligned} \text{ECM}(T_2) &= \frac{n\sigma^2}{(n + 1)^2} + \left[\mu - \frac{n\mu}{(n + 1)}\right]^2 \\ &= \frac{n\sigma^2 + \mu^2}{(n + 1)^2}. \end{aligned}$$

Si $n = 10$ y $\sigma^2 = 100$; entonces

$$\text{ECM}(T_1) = 10,$$

y

$$\text{ECM}(T_2) = (1000 + \mu^2)/121.$$

Al igualar las dos expresiones anteriores y resolver para μ, se tiene que para $\mu < \sqrt{210}$, $\text{ECM}(T_2) < \text{ECM}(T_1)$; pero si $\mu > \sqrt{210}$, entonces $\text{ECM}(T_1) < \text{ECM}(T_2)$.

Es por esta razón que se deben examinar criterios adicionales para la selección de los estimadores de θ, aun a pesar de que el error cuadrático medio sea el concepto más importante. De manera específica se estudiarán los estimadores *insesgados, consistentes, insesgado de varianza mínima* y *eficientes*. Entonces, con base en lo anterior, se presentará un concepto importante en la estimación puntual que se conoce como *estadísticas suficientes*. A lo largo de toda la discusión se supodrá la existencia de un solo parámetro desconocido. Sin embargo, debe notarse que bajo condiciones más generales estos conceptos pueden extenderse para incluir un número mayor de parámetros desconocidos.

8.2.1 Estimadores insesgados

En el error cuadrático medio de un estimador T, el término $[\theta - E(T)]$ recibe el nombre de sesgo del estimador. El sesgo de T puede ser positivo, negativo o cero. Puesto que el cuadrado del sesgo es un componente del error cuadrático medio, es razonable insistir que éste sea, en valor absoluto, lo más pequeño posible. En otras palabras, es deseable que un estimador tenga una media igual a la del parámetro que se está estimando. Lo anterior da origen a la siguiente definición.

Definición 8.2 Se dice que la estadística $T = u(X_1, X_2, \ldots, X_n)$ es un *estimador insesgado* del parámetro θ, si $E(T) = \theta$ para todos los posibles valores de θ. De esta forma, para cualquier estimador insesgado de θ, la distribución de muestreo de T se encuentra centrada alrededor de θ y $\text{ECM}(T) = Var(T)$.

En la sección 7.4 se demostró que, sin importar la distribución de la población de interés, $E(\overline{X}) = \mu$. Por lo tanto, la media muestral es un estimador insesgado de la media de la población μ para todos los valores de μ. De hecho, si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de la distribución de X con media μ, entonces cualquier X_i de la muestra un estimador insesgado de μ, dado que $E(X_i) = \mu$ para toda $i = 1, 2, \ldots, n$. Además, si una estadística T es cualquier combinación lineal de las variables aleatorias de la muestra de manera tal que

$$T = a_1X_1 + a_2X_2 + \cdots + a_nX_n$$

en donde $\sum_{i=1}^{n} a_i = 1$, entonces T es un estimador insesgado de μ dado que

$$\begin{aligned} E(T) &= E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) \\ &= a_1\mu + a_2\mu + \cdots + a_n\mu \\ &= \mu. \end{aligned}$$

En la sección 7.5 se demostró que si la varianza muestral S^2 está dada por (7.14), entonces, cuando se muestrea una distribución normal, $E(S^2) = \sigma^2$. A continuación se demostrará que si S^2 está definida por (7.14), entonces éste es un estimador insesgado de σ^2 sin importar cuál sea la distribución de la población de interés. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de alguna distribución con una función de densidad no especificada. De esta manera, $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$ para toda $i = 1, 2, \ldots, n$.

Entonces

$$\begin{aligned} E(S^2) &= E\left[\sum_{i=1}^{n} (X_i - \overline{X})^2/(n-1)\right] \\ &= (n-1)^{-1} E\left\{\sum_{i=1}^{n} [(X_i - \mu) - (\overline{X} - \mu)]^2\right\} \\ &= (n-1)^{-1} E\left\{\sum_{i=1}^{n} [(X_i - \mu)^2 - n(\overline{X} - \mu)^2]\right\}^* \\ &= (n-1)^{-1}\left[\sum_{i=1}^{n} E(X_i - \mu)^2 - nE(\overline{X} - \mu)^2\right]; \end{aligned}$$

pero por definición $E(X_i - \mu)^2 = Var(X_i) = \sigma^2$ y $E(\overline{X} - \mu)^2 = Var(\overline{X}) = \sigma^2/n$. Por lo tanto

$$\begin{aligned} E(S^2) &= (n-1)^{-1}[n\sigma^2 - (n\sigma^2)/n] \\ &= \frac{\sigma^2(n-1)}{n-1} \\ &= \sigma^2. \end{aligned}$$

En otras palabras, S^2 es un estimador insesgado de σ^2 sólo cuando el divisor es igual a $n-1$. Esta es la razón del por qué al determinar la varianza muestral se divide por $n-1$ en lugar de dividir por n. El lector debe saber que este resultado no hará de S un estimador insesgado de σ (véase la sección 11.2.2).*

8.2.2 Estimadores consistentes

Es razonable esperar que un buen estimador de un parámetro θ sea cada vez mejor conforme crece el tamaño de la muestra. Esto es, conforme la información en una muestra aleatoria se vuelve más completa, la distribución de muestreo de un buen estimador se encuentra cada vez más concentrada alrededor del parámetro θ. Se tendrá un mejor estimador de θ si se basa en 30 observaciones que si lo hace con sólo cinco. Esta idea origina lo que se conoce como un estimador consistente.

Definición 8.3 Sea T el estimador de un parámetro θ, y sea $T_1, T_2, \ldots, T_n$ una secuencia de estimadores que representan a T con base en muestras de tamaño 1, 2 ...

* Véase el material que lleva a la expresión (7.15)

n, respectivamente. Se dice que T es un estimador consistente (sencillo)* para θ si

$$\lim_{n\to\infty} P(|T_n - \theta| \leq \varepsilon) = 1$$

para todos los valores de θ y $\varepsilon > 0$.

El requisito de que $\lim_{n\to\infty} P(|T_n - \theta| \leq \varepsilon) = 1$ para toda θ constituye lo que se denomina ***convergencia en probabilidad***. Es decir, si un estimador es consistente, converge en probabilidad al valor del parámetro que está intentando estimar conforme el tamaño de la muestra crece. Esto implica que la varianza de un estimador consistente T_n disminuye conforme n crece, y la media de T_n tiende hacia donde n crece. De esta forma, las condiciones que T_n debe cumplir para ser un estimador insesgado de θ y para que $Var(T_n)\to 0$ conforme $n\to\infty$ son suficientes (pero no necesarias) para que exista consistencia. Por ejemplo, la media muestral $\overline{X}$ y la varianza muestral S^2 son estimadores consistentes de μ y σ^2, respectivamente. Para demostrar que $\overline{X}$ es un estimador consistente de μ, primero se enunciará un importante teorema conocido como desigualdad de Tchebysheff.

Teorema 8.1 Sea X una variable aleatoria con una función (densidad) de probabilidad $f(x)$ de manera tal que tanto $E(X) = \mu$ como $Var(X) = \sigma^2$ tienen un valor finito. Entonces

$$P(|X - \mu| \leq k\sigma) \geq 1 - \frac{1}{k^2}$$

o

$$P(|X - \mu| > k\sigma) \leq \frac{1}{k^2}$$

para cualquier constante $k \geq 1$. (Para la demostración de este teorema véase [3].)

La desigualdad de Tchebysheff es muy importante, ya que permite determinar los límites de las probabilidades de variables aleatorias discretas o continuas sin tener que especificar sus funciones (densidades) de probabilidad. Este teorema de Tchebysheff asegura que la probabilidad de que una variable aleatoria se aleje no más de k desviaciones estándar de la media, es menor o igual a $1/k^2$ para algún valor de $k \geq 1$. Por ejemplo

$$P(|X - \mu| \leq 2\sigma) \geq 1 - \frac{1}{4}$$

y

$$P(|X - \mu| \leq 3\sigma) \geq 1 - \frac{1}{9}$$

para cualquier variable aleatoria X con media μ y varianza σ^2 finitas.

* También puede definirse un estimador de error cuadrático consistente en forma tal que

$$\lim_{n\to\infty} E(T_n - \theta)^2 = 0, \qquad \text{para toda } \theta,$$

pero la idea de consistencia sencilla es una propiedad más básica.

Para demostrar que la media muestral $\overline{X}_n$, como función de una muestra aleatoria de tamaño n, es un estimador consistente de μ, se utilizará el resultado proporcionado por el teorema 8.1.

Teorema 8.2 Sean $X_1, X_2, \ldots, X_n$ n variables aleatorias IID, tales que $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$ tienen un valor finito para $i = 1, 2, \ldots, n$. Entonces $\overline{X}_n = \sum_{i=1}^{n} X_i/n$ es un estimador consistente de μ.

Demostración: Se quiere demostrar que

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| \leq \varepsilon) = 1.$$

Dado que $\overline{X}_n$ es una variable aleatoria tal que $E(\overline{X}_n) = \mu$ y $Var(\overline{X}_n) = \sigma^2/n$, se deduce del teorema de Tchebysheff que

$$P(|\overline{X}_n - \mu| > k\sigma/\sqrt{n}) \leq 1/k^2.$$

Sea k una constante positiva igual a $\varepsilon\sqrt{n}/\sigma$, en donde ε es un número real positivo. Entonces

$$P(|\overline{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Dado que σ^2 tiene un valor finito, tomando el límite de esta expresión conforme n tiende al infinito se tiene

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| > \varepsilon) = 0.$$

Por lo tanto, se concluye que

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| \leq \varepsilon) = 1,$$

y $\overline{X}_n$ es un estimador consistente de μ.

El teorema 8.2 también se conoce como la *ley de los grandes números*. Ésta proporciona el fundamento teórico para estimar la media de la distribución de la población con base en el promedio de un número finito de observaciones de manera tal que la confiabilidad de este promedio es mejor que la de cualquiera de las observaciones. Lo anterior permite determinar el tamaño necesario de la muestra para asegurar con determinada probabilidad que la media muestral no se alejará más allá de una cantidad específica de la media de la población.

Ejemplo 8.2 Considere el proceso de selección de una muestra aleatoria de alguna distribución que tiene una varianza conocida de $\sigma^2 = 10$ pero con una media μ desconocida. ¿Cuál debe ser el tamaño de la muestra para que la media $\overline{X}_n$ se encuentre dentro de un intervalo igual a dos unidades, de la media poblacional con una probabilidad de, por lo menos, 0.9?

Primero se desarrollará una expresión general para n. Del teorema 8.1, se sabe que

$$P(|\overline{X}_n - \mu| \leq k\sigma/\sqrt{n}) \geq 1 - \frac{1}{k^2}. \tag{8.3}$$

Elíjase un número positivo α de manera tal que $\alpha = 1/k^2$; o $k = 1/\sqrt{\alpha}$, en donde necesariamente $0 < \alpha < 1$. Entonces

$$P(|\overline{X}_n - \mu| \leq \sigma/\sqrt{n\alpha}) \geq 1 - \alpha. \tag{8.4}$$

Sea $\varepsilon > 0$ la magnitud del máximo error permisible entre $\overline{X}_n$ y μ con base en una muestra de tamaño n. Entonces

$$\varepsilon = \sigma/\sqrt{n\alpha}. \tag{8.5}$$

Resolviendo para n, se tiene

$$n = \frac{\sigma^2}{\alpha\varepsilon^2}. \tag{8.6}$$

Es claro que $\alpha = 0.1$ y $\varepsilon = 2$ para determinar los valores de n. Sustituyendo en (8.6), se tiene

$$\begin{aligned} n &= 10/(0.1)(4) \\ &= 25; \end{aligned}$$

de esta manera, si se selecciona una muestra que contenga por lo menos 25 observaciones de la distribución, el valor de la media se encontrará dentro de un intervalo con longitud de dos unidades con respecto a la media poblacional que tenga una probabilidad no menor que 0.9. El valor de probabilidad 0.9 asociado con esta afirmación en una medida de la confiabilidad con que se puede formular una inferencia respecto a μ y con base en $\overline{X}$.

8.2.3 Estimadores insesgados de varianza mínima

Para un parámetro que posee un error cuadrático medio mínimo es difícil determinar un estimador para todos los posibles valores del parámetro. Sin embargo, es posible analizar cierta clase de estimadores y dentro de esta clase intentar determinar uno que tenga un error cuadrático medio mínimo. Por ejemplo, considérese la clase de estimadores insesgados para el parámetro θ. Si una estadística T se encuentra dentro de esta clase, entonces $E(T) = \theta$ y ECM $(T) = Var(T)$. Puesto que es deseable que la varianza de un estimador sea lo más pequeña posible, debe buscarse uno en la clase de estimadores insesgados, si es que éste existe, que tenga una varianza mínima para todos los valores posibles de θ. Este estimador recibe el nombre de estimador *insesgado de varianza mínima uniforme* (VMU) de θ. La definición formal de un estimador VMU es la siguiente.

Definición 8.4 Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribucion cuya función (densidad) de probabilidad es $f(x; \theta)$. Sea la estadística $T = u(X_1, X_2, \ldots, X_n)$ un estimador de θ tal que $E(T) = \theta$ y $Var(T)$ es menor que la varianza de

cualquier otro estimador insesgado de θ para todos los posibles valores de θ. Se dice entonces que T es un estimador insesgado de varianza mínima de θ.

La varianza de un estimador insesgado es la cantidad más importante para decidir qué tan bueno es el estimador para estimar un parámetro θ. Por ejemplo, sean T_1 y T_2 cualesquiera dos estimadores insesgados de θ. Se dice que T_1 es un estimador más eficiente de θ que T_2 si $Var(T_1) \leq Var(T_2)$, cumpliéndose la desigualdad en el sentido estricto para algún valor de θ. Es muy común utilizar el cociente $Var(T_1)/Var(T_2)$ para determinar la eficiencia relativa de T_2 con respecto a T_1. Si los estimadores son sesgados, se emplean sus errores cuadráticos medios para determinar las eficiencias relativas.

¿Cómo obtener un estimador VMU, si es que éste existe? En muchos casos resulta prohibitivo determinar las varianzas de todos los estimadores insesgados de θ y entonces se selecciona el estimador que tenga la varianza más pequeña. La búsqueda de un estimador VMU se facilita bastante con la ayuda de un resultado que recibe el nombre de *cota inferior de Cramér-Rao,* el cual se presenta en el siguiente teorema. Para una demostración de éste y otros detalles que incluyen algunas condiciones de regularidad, se invita al lector a que consulte [2].

Teorema. 8.3 Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con una función (densidad) de probabilidad $f(x;\ \theta)$. Si T es un estimador insesgado de θ, entonces la varianza de T debe satisfacer la siguiente desigualdad

$$Var(T) \geq \frac{1}{nE\left[\left(\frac{\partial \ln f(X;\ \theta)}{\partial \theta}\right)^2\right]}. \tag{8.7}$$

El teorema 8.3 establece un límite inferior para la varianza de un estimador de θ. Sin embargo, lo anterior no necesariamente implica que la varianza de un estimador VMU de θ tenga que ser igual al límite inferior de Cramér-Rao. En otras palabras, es posible encontrar un estimador insesgado de θ que tenga la varianza más pequeña posible de entre todos los estimadores insesgados de θ, pero cuyas varianzas son más grandes que el límite inferior de Cramér-Rao. Un estimador de esta clase sigue siendo un estimador VMU de θ. Para un estimador insesgado cuya varianza se apega a la cota inferior de Cramér-Rao, se tiene la siguiente definición.

Definición 8.5 Si T es cualquier estimador insesgado del parámetro θ tal que

$$Var(T) = \frac{1}{nE\left[\left(\frac{\partial \ln f(X;\ \theta)}{\partial \theta}\right)^2\right]},$$

entonces se dice que T es un estimador *eficiente* de θ.

De esta forma, el estimador eficiente de θ es el estimador VMU cuya varianza es igual al límite inferior de Cramér-Rao. El estimador eficiente de θ, si es que se puede

encontrar, es el mejor estimador (insesgado) de θ en el contexto de la inferencia estadística clásica.

Ejemplo 8.3 Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución de Poisson cuya función de probabilidas es $p(x;\ \lambda) = e^{-\lambda}\lambda^x/x!$. Obtener el estimador eficiente de λ.

Dado que $p(x;\ \lambda) = \lambda^x \exp(-\lambda)/x!$,

$$\ln p(x;\ \lambda) = x\ln(\lambda) - \lambda - \ln(x!)$$

y

$$\frac{\partial \ln p(x;\ \lambda)}{\partial \lambda} = \frac{x}{\lambda} - 1$$

$$= (x - \lambda)/\lambda.$$

Entonces

$$E\left[\left(\frac{\partial \ln p(X;\ \lambda)}{\partial \lambda}\right)^2\right] = E[(X - \lambda)/\lambda]^2$$

$$= \frac{1}{\lambda^2} E(X - \lambda)^2$$

$$= \frac{Var(X)}{\lambda^2};$$

pero si X es una variable aleatoria de Poisson, $Var(X) = \lambda$. Lo anterior da como resultado

$$E\left[\left(\frac{\partial \ln p(X;\ \lambda)}{\partial \lambda}\right)^2\right] = \frac{1}{\lambda},$$

y, por la definición 8.5, la varianza del estimador eficiente de λ es

$$Var(T) = \frac{1}{n/\lambda} = \lambda/n = \sigma^2/n,$$

en donde $\sigma^2 = \lambda$ es la varianza de la población. Por lo tanto, el estimador eficiente del parámetro λ de Poisson es la media muestral $\bar{X}$.

Se concluirá esta sección sobre las propiedades deseables de los estimadores regresando al importante concepto de estadísticas suficientes. Este concepto es importante puesto que si existe un estimador eficiente, se encontrará que también es una estadística suficiente.

8.2.4 Estadísticas suficientes

De manera intuitiva, una *estadística suficiente* para un parámetro θ es aquélla que utiliza toda la información contenida en la muestra aleatoria con respecto a θ. Por

ejemplo, supóngase que $X_1, X_2, \ldots, X_{50}$ es una muestra aleatoria de 50 observaciones de una distribución gama con una función de densidad

$$f(x; 2, \theta) = \frac{1}{\theta^2} x \exp(-x/\theta) \qquad x > 0,$$

en donde el parámetro de escala θ, $\theta > 0$, es desconocido. Con una estadística suficiente para θ, lo que se tiene es una manera de resumir todas las mediciones de los datos de la muestra en un valor en el que toda la información de la muestra con respecto a θ se encuentre contenida en este valor. Para este ejemplo, el estimador

$$T = (X_1 + X_3 + \cdots + X_{49})/25$$

¿contiene toda la información pertinente con respecto a θ? A pesar de que el estimador T proporciona un solo valor, no es posible que éste contenga toda la información muestral con respecto a θ, dado que se ha excluido la mitad de las observaciones. ¿Qué puede decirse acerca de la media muestral? Con toda seguridad ésta incluye todas las observaciones de la muestra aleatoria. ¿Significa esto que toda la información muestral con respecto a θ se extrae considerando a $\bar{X}$? Se dice que una estadística $T = u(X_1, X_2, \ldots, X_n)$ es suficiente para un parámetro θ si la distribución conjunta de $X_1, X_2, \ldots, X_n$, dado T, se encuentra libre de θ; es decir, si se afirma T, entonces $X_1, X_2, \ldots, X_n$ no tiene nada más qué decir con respecto a θ.

La utilidad de una estadística suficiente recae en el hecho de qué si un estimador insesgado de un parámetro θ es una función de una estadística suficiente, entonces tendrá la varianza más pequeña de entre todos los estimadores insesgados de θ que no se encuentren basados en una estadística suficiente. De hecho, si existe el estimador eficiente de θ, se encontrará que éste es una estadística suficiente. Un criterio para determinar una estadística suficiente está dado por el siguiente teorema, el cual se conoce como *teorema de factorización de Neyman.*

Teorema 8.4 Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con una función de densidad de probabilidad $f(x; \theta)$. Se dice que la estadística $T = u(X_1, X_2, \ldots, X_n)$ es una estadística suficiente para θ si y sólo si la función de verosimilitud puede factorizarse de la siguiente forma:

$$L(x_1, x_2, \ldots, x_n; \theta) = h(t; \theta)\, g(x_1, x_2, \ldots, x_n)$$

para cualquier valor $t = u(x_1, x_2, \ldots, x_n)$ de T y en donde $g(x_1, x_2, \ldots, x_n)$ no contiene al parámetro θ.

Ejemplo 8.4 Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución gama cuya función de densidad de probabilidad es

$$f(x; \theta) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} \exp(-x/\theta) \qquad x > 0,$$

y en donde el valor del parámetro de forma α es conocido. Obtener una estadística suficiente para el parámetro de escala θ.

La función de verosimilitud es

$$\begin{aligned}
L(x_1, x_2, \ldots, x_n; \theta) &= f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta) \\
&= \frac{1}{\Gamma(\alpha)\theta^\alpha} x_1^{\alpha-1} \exp(-x_1/\theta) \cdot \frac{1}{\Gamma(\alpha)\theta^\alpha} x_2^{\alpha-1} \exp(-x_2/\theta) \\
&\quad \cdots \frac{1}{\Gamma(\alpha)\theta^\alpha} x_n^{\alpha-1} \exp(-x_n/\theta) \\
&= \frac{1}{\Gamma^n(\alpha)\theta^{n\alpha}} \prod_{i=1}^{n} x_i^{\alpha-1} \exp\left(-\sum_{i=1}^{n} x_i/\theta\right) \\
&= \frac{1}{\theta^{n\alpha}} \exp\left(-\sum x_i/\theta\right) \cdot \frac{\Pi x_i^{\alpha-1}}{\Gamma^n(\alpha)} \\
&= h\left(\sum x_i; \theta\right) g(x_1, x_2, \ldots, x_n).
\end{aligned}$$

Por el teorema 8.4, $\Sigma_{i=1}^{n} X_i$ es una estadística suficiente para θ.

Supóngase, en el ejemplo 8.4, que se considera un estimador de θ de la forma

$$T = \frac{1}{n\alpha} \sum_{i=1}^{n} X_i. \tag{8.8}$$

puede verse que T es una función de la estadística suficiente ΣX_i.
Por lo tanto, T también es una estadística suficiente para θ dado que la función de verosimilitud para el ejemplo 8.4, puede factorizarse como

$$L(x_1, x_2, \ldots, x_n) = h(t; \theta)\, g(x_1, x_2, \ldots, x_n).$$

en donde $\Sigma X_i = n\alpha T$ y

$$h(t; \theta) = \frac{1}{\theta^{n\alpha}} \exp(-n\alpha t/\theta). \tag{8.9}$$

Como resultado se tiene que se satisfacen las condiciones del teorema de factorización. De hecho, puede demostrarse que cualquier función uno a uno de una estadística suficiente, también es suficiente.

Ejemplo 8.5 Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución de Poisson cuya función de probabilidad es

$$p(x; \lambda) = \lambda^x \exp(-\lambda)/x! \qquad x = 0, 1, 2, \ldots .$$

Demostrar que el estimador eficiente de λ es a su vez una estadística suficiente.

Del ejemplo 8.3, recuérdese que el estimador eficiente de λ es la media muestral $\bar{X}$. Se necesita demostrar que λ es una función uno a uno de una estadística suficiente para λ. La función de verosimilitud es

$$\begin{aligned}L(x_1, x_2, \ldots, x_n; \lambda) &= p(x_1; \lambda)\, p(x_2; \lambda) \cdots p(x_n; \lambda)\\ &= \frac{\lambda^{x_1}\exp(-\lambda)}{x_1!} \cdot \frac{\lambda^{x_2}\exp(-\lambda)}{x_2!} \cdots \frac{\lambda^{x_n}\exp(-\lambda)}{x_n!}\\ &= \lambda^{\sum_{i=1}^{n} x_i}\exp(-n\lambda)/\prod_{i=1}^{n} x_i!\\ &= h(\Sigma x_i; \lambda)\, g(x_1, x_2, \ldots, x_n)\end{aligned}$$

en donde

$$h(\Sigma x_i; \lambda) = \lambda^{\Sigma x_i}\exp(-n\lambda).$$

Por el teorema 8.4, la estadística $\Sigma_{i=1}^{n} X_i$ es suficiente para λ. Dado que el estimador $\bar{X}$ es una función uno a uno de esta estadística, $\bar{X}$ también es suficiente para λ.

8.3 Métodos de estimación puntual

En la sección anterior se mencionaron las propiedades deseables de un buen estimador. En esta sección se estudiará cómo obtener estimadores que, de manera general, tengan buenas propiedades. Específicamente se considerarán los métodos de máxima verosimilitud y el de momentos. En el capítulo 13 se encontrará el método de mínimos cuadrados que se emplea para ajustar ecuaciones.

8.3.1 Estimación por máxima verosimilitud

Para introducir el concepto de estimación de máxima verosimilitud, piense en el siguiente hecho. El desborde de ríos y lagos es un fenómeno natural que a veces tiene devastadoras consecuencias. Supóngase que en cierto año hubo dos serias inundaciones, por este fenómeno, en determinada región geográfica. Si se supone que el número de inundaciones por año en esta localidad es una variable aleatoria de Poisson con un valor del parámetro λ, desconocido, ¿cómo debe procederse para estimar el valor de λ con base en una sola observación $x = 2$? Un posible método es seleccionar el valor de λ para el cual la probabilidad del valor observado es máxima. Es posible, para el valor observado, que λ sea cualquier número positivo. Para propósitos de la presentación, supóngase que los posibles valores de λ son 1, 3/2, 2, 5/2 y 3. Las probabilidades para el valor observado $x = 2$ para cada uno de estos valores de λ son las siguientes:

λ	1	3/2	2	5/2	3
$p(2; \lambda)$	0.1839	0.2510	0.2707	0.2565	0.2240

Aparentemente $p(2; \lambda)$ crece hasta un valor máximo de 0.2707 para $\lambda = 2$, y disminuye para $\lambda > 2$. El valor de 2 de λ es el que maximiza la probabilidad del valor observado. En otras palabras, la observación $x = 2$ tiene una probabilidad mayor de ocurrencia para una distribución de Poisson con $\lambda = 2$ que para cualquier

otro valor del parámetro λ. Puede demostrarse que el valor $\lambda = 1$ es el que maximiza a $\lambda = 2$ tomando la primera derivada de $p(2; \lambda)$ con respecto a λ e igualándola a cero. Dado que

$$p(2; \lambda) = \lambda^2 \exp(-\lambda)/2!,$$

se tiene

$$\frac{dp(2; \lambda)}{d\lambda} = \frac{1}{2}[-\lambda^2 \exp(-\lambda) + 2\lambda \exp(-\lambda)]$$

$$= \frac{\lambda \exp(-\lambda)}{2}(2-\lambda).$$

Igualando la primera derivada a cero se tienen las raíces $\lambda = 0$ o $\lambda = 2$. La segunda derivada con respecto a λ da como resultado la expresión $\exp(-\lambda)[1 - 2\lambda + (\lambda^2)/2]$, cuyo valor para $\lambda = 2$ es $-\exp(-2) < 0$. De esta forma, el valor $x = 2$ es aquél para el cual el valor de la probabilidad de la observación es máximo. Este valor recibe el nombre de estimador de máxima verosimilitud.

En esencia, el método de estimación por máxima verosimilitud, selecciona como estimador a aquél valor del parámetro que tiene la propiedad de maximizar el valor de la probabilidad de la muestra aleatoria observada. En otras palabras, el método de máxima verosimilitud consiste en encontrar el valor del parámetro que maximiza la función de verosimilitud.

Definición 8.6 Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con función (densidad) de probabilidad $f(x; \theta)$, y sea $L(x_1, x_2, \ldots, x_n; \theta)$ la verosimilitud de la muestra como función de θ. Si $t = u(x_1, x_2, \ldots, x_n)$ es el valor de θ para el cual el valor de la función de verosimilitud es máxima, entonces $T = u(X_1, X_2, \ldots, X_n)$ es el estimador de máxima verosimilitud de θ, y t es el *estimador de máxima verosimilitud.*

El método de máxima verosimilitud (MV) tiene la propiedad (deseable) de proporcionar estimadores que son funciones de estadísticas suficientes, siempre y cuando el estimador MV sea único. Además, el método MV proporciona el estimador eficiente, si es que existe. Sin embargo, los estimadores MV son generalmente sesgados. El procedimiento para obtener este tipo de estimadores es (relativamente) directo. Debido a la naturaleza de la función de verosimilitud se escoge, por lo común, maximizar el logaritmo natural de $L(\theta)$. Esto es, en muchas ocasiones es más fácil obtener el estimado MV maximizando $\ln L(\theta)$ que $L(\theta)$. En los siguientes ejemplos se ilustra el método.

Ejemplo 8.6 En un experimento binomial se observan $X = x$ éxitos en n ensayos. Obtener el estimador de máxima verosimilitud del parámetro binomial p.

En este caso la función de verosimilitud es idéntica a la probabilidad de que $X = x$; de esta forma

$$L(x; p) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}, \qquad 0 \leq p \leq 1.$$

Entonces

$$\ln L(x;\, p) = \ln(n!) - \ln[(n - x)!] - \ln(x!) + x\ln(p) + (n - x)\ln(1 - p).$$

Para encontrar el valor de p, para el cual $\ln L\,(x;\, p)$ tiene un valor máximo, se toma la primera derivada con respecto a p y se iguala a cero:

$$\frac{d[\ln L(x;\, p)]}{dp} = \frac{x}{p} - \frac{(n - x)}{(1 - p)} = 0.$$

Después de resolver para p, se obtiene el estimador MV de p el cual recibe el nombre de proporción muestral X/n, y el estimado MV es x/n. Para confirmar que este valor maximiza a $\ln L(x;\, p)$, se toma la segunda derivada con respecto a p y se evalúa en x/n:

$$\frac{d^2[\ln L(x;\, p)]}{dp^2} = -\frac{np(1 - p) + (x - np)(1 - 2p)}{[p(1 - p)]^2}$$

y

$$\left.\frac{d^2\,[\ln L(x;\, p)]}{dp^2}\right|_{x/n} = -\frac{x}{(x/n)^2[1 - (x/n)]} \qquad x/n < 1,$$

lo que confirma el resultado, dado que la segunda derivada es negativa. Para un ejemplo específico, si se observan $x = 5$ con base en 25 ensayos independientes, el estimado MV de p es $5/25 = 0.2$.

Ejemplo 8.7 Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución normal con una función de densidad de probabilidad

$$f(x;\, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp[-(x - \mu)^2/2\sigma^2].$$

Determinar los estimadores de μ y σ^2.

Para este problema se procederá de la misma forma que en el caso de un solo parámetro. Dado que la función de verosimilitud depende tanto de μ como de σ^2, los estimados MV de μ y σ^2 son los valores para los cuales la función de verosimilitud tiene un valor máximo. De acuerdo con lo anterior

$$L(x_1, x_2, \ldots, x_n;\, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp[-(x_1 - \mu)^2/2\sigma^2] \cdots \frac{1}{\sqrt{2\pi}\sigma} \times \exp[-(x_n - \mu)^2/2\sigma^2]$$

$$= (2\pi\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right].$$

y

$$\ln L(x_1, x_2, \ldots, x_n;\, \mu, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Después de obtener las primeras derivadas parciales con respecto a μ y con respecto a σ^2 e igualándolas a cero, se tiene

$$\frac{\partial[\ln L(\mu, \sigma^2)]}{\partial\mu} = -\frac{2}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

y

$$\frac{\partial[\ln L(\mu,\sigma^2)]}{\partial(\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0.$$

Resolviendo la primera ecuación para μ, sustituyendo este valor en la segunda y resolviendo para σ^2, se tiene

$$\hat{\mu} = \sum_{i=1}^{n} x_i/n = \bar{x}$$

y

$$\hat{\sigma}^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/n.$$

A pesar de que no se verificará que estos valores maximizan la función de verosimilitud, ellos son los estimados MV de μ y σ^2, respectivamente. Si existe alguna duda tómense las segundas derivadas. Sin embargo, dado que una función de verosimilitud es el producto, ya sea de probabilidades o de densidades, éstas generalmente se encuentran acotadas y son continuas en los parámetros. En consecuencia, el resultado usual es que la solución de la primera derivada proporcionará el valor para el cual la función es máxima.

Nótese que se ha introducido la acostumbrada notación "sombrero" $\hat{}$ para denotar un estimador MV. Se empleará esta notación cuando sea necesario. Nótese también que el estimador MV de σ^2 es sesgado, confirmándose de esta manera un comentario anterior en el sentido en el que los estimadores MV no necesariamente son insesgados.

El método de máxima verosimilitud posee otra propiedad deseable conocida como propiedad de invarianza. Sea $\hat{\theta} = u(X_1, X_2, \ldots, X_n)$ el estimador de máxima verosimilitud de θ. Si $g(\theta)$ es una función univaluada de θ, entonces el estimador de máxima verosimilitud de $g(\theta)$ es $g(\hat{\theta})$. Por ejemplo, dado que, cuando se muestrea una distribución normal, el estimador MV de σ^2 es

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

por la propiedad de invarianza, el estimador MV de la desviación estándar σ es

$$\hat{\sigma} = \left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{1/2}$$

Como ejemplo adicional de la propiedad de invarianza, el estimador MV de la función de confiabilidad Weibull es

$$\hat{R}(t) = \exp[-(t/\hat{\theta})^{\alpha}],$$

en donde $\hat{\theta}$ es el estimador MV del parámetro de escala θ.

8.3.2 Método de los momentos

Quizá el método más antiguo para la estimación de parámetros es el método de los momentos. Éste consiste en igualar los momentos apropiados de la distribución de la población con los correspondientes momentos muestrales para estimar un parámetro desconocido de la distribución.

Definición 8.7 Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con función (densidad) de probabilidad $f(x;\ \theta)$. El r-ésimo momento alrededor del cero se define como

$$M'_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r.$$

El método de los momentos proporciona una alternativa razonable cuando no se pueden determinar los estimadores de máxima verosimilitud. Recuérdese que los parámetros son, en general, funciones de los momentos teóricos. Por ejemplo, si la variable aleatoria X tiene una distribución gama (véase la sección 5.5), entonces

$$\mu = \alpha\theta \tag{8.10}$$

y

$$\mu'_2 = \alpha(\alpha + 1)\theta^2. \tag{8.11}$$

Resolviendo (8.10) para α y sustituyendo en (8.11), se tiene

$$\alpha = \mu/\theta \tag{8.12}$$

y

$$\mu'_2 = \frac{\mu}{\theta}\left(\frac{\mu}{\theta} + 1\right)\theta^2$$
$$= \mu^2 + \mu\theta,$$

o

$$\theta = (\mu'_2 - \mu^2)/\mu. \tag{8.13}$$

Sustituyendo (8.13) para θ en (8.12), se obtiene

$$\alpha = \mu^2/(\mu'_2 - \mu^2). \tag{8.14}$$

De esta forma, los dos parámetros de la distribución gama son funciones de los primeros dos momentos alrededor del cero.

En esencia, el método se implementa igualando tantos momentos muestrales con los correspondientes momentos teóricos tantas veces como sea necesario para determinar un estimador de momentos para un parámetro desconocido. Por ejemplo, por (8.13) y (8.14), los estimadores de momento de los parámetros gama θ y α son

$$\tilde{\theta} = (M_2' - \bar{X}^2)/\bar{X} \tag{8.15}$$

y

$$\tilde{\alpha} = \bar{X}^2/(M_2' - \bar{X}^2), \tag{8.16}$$

respectivamente, en donde se emplea la notación de tilde ($\tilde{}$) para denotar un estimador de momentos. Como ilustración adicional, recuérdese el ejemplo 4.10. Se demostrará que los parámetros p y k de una distribución binomial negativa también son funciones de los primeros dos momentos alrededor del cero, ya que

$$p = \mu/(\mu_2' - \mu^2)$$

y

$$k = \mu^2/(\mu_2' - \mu^2 - \mu).$$

Por lo tanto, los estimadores de momentos de p y k están dados por

$$\tilde{p} = \bar{X}/(M_2' - \bar{X}^2) \tag{8.17}$$

y

$$\tilde{k} = \bar{X}^2/(M_2' - \bar{X}^2 - \bar{X}), \tag{8.18}$$

respectivamente.

8.3.3 Estimación por máxima verosimilitud para muestras censuradas

En algunas situaciones de muestreo, en forma especial en las pruebas de duración, el procedimiento de prueba puede terminar antes de proporcionar una muestra aleatoria completa. En esta sección se considerará el principio de máxima verosimilitud para la estimación de parámetros desconocidos con base en este tipo de muestras, las cuales reciben el nombre de muestras *censuradas* o *truncadas*. En este contexto, las ideas se concentrarán, en forma exclusiva, alrededor de la noción de una prueba de duración.

Una prueba típica de duración consiste en artículos iguales (tales como componentes eléctricos o mecánicos) seleccionados en forma aleatoria de un proceso y operados en un medio cuidadosamente controlado hasta que el artículo falla. En este caso, la medición de interés es el lapso de tiempo que cada unidad tarda en fallar. Si la prueba de duración se termina sólo cuando todas las unidades de la muestra han fallado, se dice que la muestra aleatoria de tiempos está completa. Sin embargo, por restricciones económicas y de tiempo, generalmente la prueba termina ya sea después de un lapso de tiempo predeterminado x_0 o después de que falla un determinado número de unidades $m \leq n$. Las dos condiciones producen muestras censura-

das. Si X_0 es un lapso fijo de tiempo, el número de unidades que fallan de las n, desde el comienzo de la prueba hasta el tiempo x_0, es una variable aleatoria; ésta constituye una muestra censurada de tipo I. Si m es fijo y el tiempo de terminación X_m es la variable aleatoria, se dice que la muestra es de tipo II. Sin considerar la inferencia, existe muy poca diferencia entre estos dos tipos de muestras. De acuerdo con lo anterior, se restringirá la presentación al muestreo censurado de tipo II.

Los datos muestrales de una prueba de duración son los tiempos en los que se dio una falla. Por ejemplo, supóngase que la primera falla ocurrió en un tiempo igual a x_1 desde el comienzo, la segunda se presenta a x_2 desde el comienzo y así hasta que ocurre la m-ésima falla en un tiempo por x_m, en donde $m \leqslant n$ es el número, fijado de antemano, necesario para terminar la prueba. Los tiempos que se observaron de falla $x_1, x_2, \ldots, x_m$ constituyen una secuencia ordenada, porque $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_m$. Nótese que en el momento en que se da por terminada la prueba, existen $n - m$ unidades que todavía no han fallado; estas $n - m$ unidades tienen un tiempo de supervivencia x_m. Es claro que se tiene el tamaño completo de la muestra cuando $m = n$.

Supóngase que los tiempos de duración de las unidades son variables aleatorias $X_1, X_2, \ldots, X_n$ independientes exponencialmente distribuidas, con una función de densidad

$$f(x; \theta) = \frac{1}{\theta}\exp(-x/\theta), \qquad x > 0, \quad \theta > 0.$$

El interés recae en encontrar el estimador de máxima verosimilitud del parámetro θ. La función de verosimilitud para un muestreo censurado del tipo II es la probabilidad conjunta de que fallen m unidades en los tiempos $x_1, x_2, \ldots, x_m$ en ese orden, y sobrevivan $n - m$ unidades con un tiempo de supervivencia igual a x_m. La parte de la función de verosimilitud que corresponde a las m unidades que han fallado en los tiempos $x_1, x_2, \ldots, x_m$, es $f(x_1; \theta)f(x_2; \theta) \cdots f(x_m; \theta)$. Pero ésta es sólo una de las posibles formas en que pueden fallar m unidades de un total de n. El número total de formas es $n!/(n - m)!$. La probabilidad de que $n - m$ unidades sobrevivan un tiempo x_m, está dada por la función de confiabilidad a tiempo x_m; de esta forma, para la distribución exponencial,

$$P(X > x_m) = \exp(-x_m/\theta).$$

Por lo tanto, la función de verosimilitud es

$$L(x_1, x_2, \ldots, x_m; \theta) = \frac{n!}{(n-m)!}\Bigg\{\underbrace{\frac{1}{\theta}\exp(-x_1/\theta)\cdots\frac{1}{\theta}\exp(-x_m/\theta)}_{m \text{ términos}}\;\underbrace{\exp(-x_m/\theta)\cdots\exp(-x_m/\theta)}_{(n-m)\text{ términos}}\Bigg\}$$

$$= \frac{n!}{(n-m)!}\left\{\frac{1}{\theta^m}\exp\left(-\frac{1}{\theta}\sum_{i=1}^{m} x_i\right)\cdot\exp\left[-\frac{(n-m)}{\theta}x_m\right]\right\}$$

$$= \frac{n!}{(n-m)!}\left[\frac{1}{\theta^m}\exp\left(-\frac{1}{\theta}T_m\right)\right], \tag{8.19}$$

en donde

$$T_m = \sum_{i=1}^{m} x_i + (n - m)x_m. \tag{8.20}$$

Tomando el logaritmo natural de L, se tiene

$$\ln L(x_1, x_2, \ldots, x_m; \theta) = \ln(n!) - \ln[(n - m)!] - m\ln\theta - \frac{1}{\theta}T_m.$$

Entonces

$$\frac{d[\ln L(x_1, x_2, \ldots, x_m; \theta)]}{d\theta} = -\frac{m}{\theta} + \frac{1}{\theta^2}T_m,$$

e igualando la derivada a cero, el estimado de máxima verosimilitud de θ es

$$\hat{\theta} = \left[\sum_{i=1}^{m} x_i + (n - m)x_m\right] \Big/ m. \tag{8.21}$$

Ejemplo 8.8 Las calculadoras científicas de bolsillo comúnmente disponibles contienen paquetes de batería que deben reemplazarse después de una cierta cantidad de tiempo de uso. Supóngase que de un proceso de producción se seleccionan, en forma aleatoria, 50 paquetes de baterías y se someten a una prueba de duración. Se decide terminar la prueba cuando 15 de los 50 dejan de funcionar de manera adecuada. Los tiempos observados, en orden, en los que ocurrió la falla, son 115, 119, 131, 138, 142, 147, 148, 155, 158, 159, 163, 166, 167, 170 y 172. Si los anteriores valores son realizaciones de un conjunto de variables aleatorias independientes exponencialmente distribuidas, se debe obtener el estimado de máxima verosimilitud para θ.

En este ejemplo,

$$n = 50,\ m = 15,\ \sum_{i=1}^{15} x_i = 115 + 119 + \cdots + 172 = 2250, \quad \text{y} \quad x_{15} = 172.$$

Por lo tanto, por (8.21),

$$\hat{\theta} = \frac{2250 + (50 - 15)172}{15} = 551.33 \text{ horas.}$$

8.4. Estimación por intervalo

Para introducir la noción de una estimación por intervalo, supóngase que una tienda mantiene muy buenos registros con respecto al número de unidades de cierto producto que vende mensualmente. Para la compañía es muy importante conocer la demanda promedio ya que con base en ésta se lleva a cabo el mantenimiento del invent[illegible] S[illegible]pó[illegible] q[illegible] la demanda del producto no se ve afectada por fluctuaciones

la media muestral es $\bar{x} = 200$ unidades. En otras palabras, $\bar{x} = 200$ es un estimado puntual de un parámetro desconocido, el cual representa la demanda promedio de este producto en la tienda. Este estimador, ¿implica que la demanda media desconocida no sea mayor de 250 ni menor de 150? En este punto no es posible saberlo, ya que no se tiene ninguna indicación del posible error en el estimado puntual. El error en el estimado puntual se mide en términos de la variación muestral del correspondiente estimador.

Por ejemplo, supóngase que la desviación estándar de la media muestral $\bar{X}$ es 60 unidades. De acuerdo con el teorema central del límite, puede argumentarse que $\bar{X} \rightarrow N(\mu, 60)$, conforme $n \rightarrow \infty$. De esta forma, la probabilidad de que $\bar{X}$ se encuentre dentro de dos desviaciones estándar alrededor de μ, es de, aproximadamente, 0.95. En otras palabras, para n grande,

$$P(|\bar{X} - \mu| < 120) = 0.95,$$

o

$$P(-120 < \bar{X} - \mu < 120) = 0.95. \tag{8.22}$$

Restando $\bar{X}$ y multiplicando por -1 en el interior de los paréntesis, se tiene

$$P(\bar{X} - 120 < \mu < \bar{X} + 120) = 0.95. \tag{8.23}$$

Si se sustituye el estimado para $\bar{x} = 200 \bar{X}$, se tiene

$$P(80 < \mu < 320) = 0.95, \tag{8.24}$$

lo que sugiere que es enteramente posible que la demanda sea tan grande como 250 unidades o tan pequeña como 150 unidades, siempre que $d.e.(\bar{X}) = 60$. Por otro lado, supóngase que la desviación estándar de $\bar{X}$ es igual a 10. Entonces, la expresión correspondiente a (8.23), es

$$P(\bar{X} - 20 < \mu < \bar{X} + 20) = 0.95,$$

y para $\bar{x} = 200$,

$$P(180 < \mu < 220) = 0.95.$$

En este caso es poco probable que μ sea tan grande como 250 o tan pequeño como 150.

En ambos casos la clave para resolver el problema se encuentra en la desviación estándar del estimador puntual. En esencia, para la estimación del intervalo se consideran, tanto el estimador puntual del parámetro θ, como su distribución de muestreo, con el propósito de determinar un intervalo que, con cierta seguridad, contiene a θ.

Para tener una mayor idea acerca de la estimación por intervalo, es necesario interpretar el significado de (8.23) y (8.24). Dado que $\bar{X}$ es una variable aleatoria, el intervalo $\bar{X} - 120$ a $\bar{X} + 120$ es un intervalo aleatorio, y la probabilidad de que este intervalo contenga el valor verdadero de μ es de 0.95. En otras palabras, si se obtuviesen muestras del mismo tamaño en forma repetida de una población, y cada vez que éstas se seleccionan, se calculan los valores específicos para el intervalo aleatorio $(\bar{X} - 120, \bar{X} + 120)$; entonces debe esperarse que un 95% de estos intervalos

contengan el valor de la media desconocida μ. Por otro lado, el intervalo específico entre 80 y 320 no es más que una realización del intervalo aleatorio $(\overline{X} - 120, \overline{X} + 120)$, con base en los datos de una sola muestra, en la que el estimado es $\bar{x} = 200$. Dado que el valor de probabilidad de 0.95 se refiere sólo al intervalo aleatorio $(\overline{X} - 120, \overline{X} + 120)$, es incorrecto decir que la probabilidad de que μ se encuentre contenido en el intervalo (80, 320) es de 0.95. Esto es, no puede asociarse ningún valor de probabilidad a la proposición $80 < \mu < 320$, debido a que ésta contiene sólo constantes. Sin embargo, la probabilidad de 0.95 para el intervalo aleatorio sugiere que la confianza en que el intervalo (80, 320) contenga el valor de la media desconocida μ es alta. Sólo en este sentido se permite asignar un grado de confianza a la proposición $80 < \mu < 320$ igual a la probabilidad del intervalo aleatorio $(\overline{X} - 120, \overline{X} + 120)$; así, cuando se escribe

$$P(80 < \mu < 320) = 0.95,$$

no se está formulando ninguna proposición probabilística en el sentido clásico, sino más bien se expresa un grado de confianza. De acuerdo con lo anterior, el intervalo (80, 320) recibe el nombre de intervalo de confianza del 95% para μ.

En términos generales, la construcción de un intervalo de confianza para un parámetro desconocido θ consiste en encontrar una estadística suficiente T y relacionarla con otra variable aleatoria $X^* = f(T; \theta)$, en donde X involucra a θ pero la distribución de X no contiene a θ, así como tampoco a ningún otro parámetro desconocido. Entonces se seleccionan dos valores x_1 y x_2 tales que

$$P(x_1 < X < x_2) = 1 - \alpha,$$

en donde $1 - \alpha$ recibe el nombre de *coeficiente de confiaza*. Mediante una manipulación algebraica de las dos expresiones, se puede modificar el contenido entre paréntesis y expresarlo como

$$P[h_1(T) < \theta < h_2(T)] = 1 - \alpha,$$

en donde $h_1(T)$ y $h_2(T)$ son funciones de la estadística T y de esta forma, variables aleatorias. El intervalo de confianza para θ se obtiene sustituyendo en $h_1(T)$ y $h_2(T)$ los estimadores calculados a partir de los datos muestrales, dando origen a lo que se conoce como intervalo de confianza bilateral. Al seguirse el mismo procedimiento, también pueden desarrollarse intervalos de confianza unilaterales, de la forma

$$P[g_1(T) < \theta] = 1 - \alpha$$

o

$$P[\theta < g_2(T)] = 1 - \alpha.$$

El primero es un intervalo de confianza unilateral inferior para θ, y el segundo es un intervalo de confianza unilateral superior.

A continuación se examinarán varias situaciones que involucran la construcción de intervalos de confianza para medias y varianzas poblacionales. Será aparente que

* Este método recibe, en general, el nombre de *método pivotal*, y X se conoce entonces como *variable aleatoria pivotal*.

la discusión aquí presentada tiene un fuerte parecido al material de las secciones 7.4 a 7.8.

8.4.1 Intervalos de confianza para μ cuando se muestrea una distribución normal con varianza conocida

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución normal con media desconocida μ, pero con una varianza σ^2 conocida. El interés recae en la construcción de un intervalo de confianza de un $100(1 - \alpha)\%$ sobre μ y en donde α es un número pequeño, tal que $0 < \alpha < 1$. La construcción de un intervalo de confianza se hace con base en el mejor estimador de μ, explícitamente la media muestral $\overline{X}$.

Para ilustrar el enfoque fundamental para la construcción de intervalos de confianza, considérese la proposición probabilística dada por (8.22). Sumando μ dentro de los paréntesis, se tiene

$$P(\mu - 120 < \overline{X} < \mu + 120) = 0.95.$$

De esta forma, los límites $\mu - 120$ y $\mu + 120$ son funciones de los posibles valores de μ. Por lo tanto, y en general, se puede escribir

$$P[g_1(\mu) < \overline{X} < g_2(\mu)] = 1 - \alpha, \tag{8.25}$$

de manera tal que

$$\int_{-\infty}^{g_1(\mu)} f(\bar{x}; \mu)d\bar{x} = \alpha/2$$

y

$$\int_{g_2(\mu)}^{\infty} f(\bar{x}; \mu)d\bar{x} = \alpha/2,$$

en donde $f(\bar{x}; \mu)$ es la función de densidad de la distribución de muestreo de $\overline{X}$, y $g_1(\mu)$ y $g_2(\mu)$ son funciones de μ las cuales no contienen a ningún otro parámetro desconocido.

De interés inmediato es la determinación de $g_1(\mu)$ y $g_2(\mu)$. Dado que $\overline{X} \sim N(\mu, \sigma/\sqrt{n})$, la normal estándar $Z = (\overline{X} - \mu)/(\sigma/\sqrt{n})$, y

$$P[g_1(\mu) < \overline{X} < g_2(\mu)] = P\left[\frac{g_1(\mu) - \mu}{\sigma/\sqrt{n}} < Z < \frac{g_2(\mu) - \mu}{\sigma/\sqrt{n}}\right] = 1 - \alpha. \tag{8.26}$$

Pero ya que $P(z_{\alpha/2} < Z < z_{1-\alpha/2}) = 1 - \alpha$, en donde los valores cuantiles $z_{\alpha/2}$ y $z_{1-\alpha/2}$ son tales que $P(Z < z_{\alpha/2}) = \alpha/2$ y $P(Z < z_{1-\alpha/2}) = 1 - \alpha/2$, respectivamente, se sigue que

$$\frac{g_1(\mu) - \mu}{\sigma/\sqrt{n}} = z_{\alpha/2} \tag{8.27}$$

y

$$\frac{g_2(\mu) - \mu}{\sigma/\sqrt{n}} = z_{1-\alpha/2}. \tag{8.28}$$

Dando solución a (8.27) y (8.28) en términos de $g_1(\mu)$ y $g_2(\mu)$, respectivamente, se obtienen

$$g_1(\mu) = \mu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{8.29}$$

y

$$g_2(\mu) = \mu + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}. \tag{8.30}$$

Dado que para la normal estándar $z_{\alpha/2} = -z_{1-\alpha/2}$, puede sustituirse $-z_{1-\alpha/2}$ para $z_{\alpha/2}$ en (8.29). De acuerdo con lo anterior, pueden sustituirse las expresiones (8.29) y (8.30) para $g_1(\mu)$ y $g_2(\mu)$, respectivamente, en (8.25) para obtener

$$P\left(\mu - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{X} < \mu + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha. \tag{8.31}$$

Al manipular las desigualdades que se encuentran dentro de los paréntesis en (8.31), se tiene

$$P\left(\overline{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha, \tag{8.32}$$

que es una generalización de la proposición probabilística (8.23). La probabilidad de que el intervalo aleatorio de $\overline{X} - z_{1-\alpha/2}(\sigma/\sqrt{n})$ a $\overline{X} + z_{1-\alpha/2}(\sigma/\sqrt{n})$ contenga el verdadero valor de la media μ es $1 - \alpha$. Si se reemplaza la variable aleatoria X en (8.32) por el estimado $\bar{x}$ calculado a partir de los datos de una muestra de tamaño n, un intervalo de confianza del $100(1 - \alpha)\%$ para μ, es

$$\bar{x} \pm z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \tag{8.33}$$

en donde $\bar{x} - z_{1-\alpha/2}(\sigma/\sqrt{n})$ y $\bar{x} + z_{1-\alpha/2}(\sigma/\sqrt{n})$ reciben el nombre de límites de confianza inferiores y superiores, respectivamente, para μ. Esto es, el intervalo de confianza (8.33) es un intervalo estimado para μ.

Al examinar el intervalo de confianza para μ dado por (8.33), es fácil, relativamente, observar que entre más grande es el tamaño de la muestra, más pequeño es el ancho del intervalo; o para un coeficiente de confianza $1 - \alpha$ más grande, mayor es el ancho del intervalo. Ambos resultados son lógicos ya que un tamaño grande de la muestra disminuirá la varianza del estimador, y un coeficiente de confianza grande incrementa el valor cuantil dando como resultado un intervalo más amplio.

Ejemplo 8.9 Los datos que a continuación se dan son los pesos en gramos del contenido de 16 cajas de cereal que se seleccionaron de un proceso de llenado con el propósito de verificar el peso promedio: 506, 508, 499, 503, 504, 510, 497, 512, 514, 505, 493, 496, 506, 502, 509, 496. Si el peso de cada caja es una variable aleatoria normal con una desviación estándar $\sigma = 5$ g, obtener los intervalos de confianza estimados del 90, 95 y 99%, para la media de llenado de este proceso.

Para un coeficiente de confianza del 90%, $\alpha = 0.1$. El valor $z_{0.95}$ se obtiene de la tabla D del apéndice y es igual a 1.645, ya que $P(Z > 1.645) = 0.05$. Con base en los datos muestrales, el valor de $\bar{x}$ es de 503.75 g. Entonces un intervalo de confianza del 90% para la media del proceso de llenado es

$$503.75 \pm 1.645 \frac{5}{\sqrt{16}},$$

o de 501.69 a 505. 81. Los otros intervalos de confianza deseados se obtienen siguiendo el mismo procedimiento. Los resultados se encuentran resumidos en la tabla 8.1.

En este momento se considerará un problema que es enteramente similar al del ejemplo 8.2. Supóngase que se especifica que el muestreo se efectúa sobre una población que tiene una distribución normal con media μ desconocida y varianza σ^2 conocida. Se desea estimar el tamaño necesario de la muestra de manera tal que, con una probabilidad de $1 = \alpha$, la media muestral $\bar{X}$ se encuentre en un intervalo igual a ε unidades alrededor de la media de la población μ. La expresión (8.31) puede reescribirse como

$$P\left(-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha, \qquad (8.34)$$

la cual da como resultado

$$P(|\bar{X} - \mu| < \varepsilon) = 1 - \alpha$$

en donde

$$\varepsilon = z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}. \qquad (8.35)$$

Al resolver para n en (8.35) se obtiene el resultado deseado,

$$n = \left(\frac{z_{1-\alpha/2}\sigma}{\varepsilon}\right)^2 \qquad (8.36)$$

La única diferencia entre las expresiones (8.6) y (8.36) es que la primera se obtuvo sin especificar la distribución de la población, mientras que para la segunda se supuso que el muestreo se llevaba a cabo sobre una distribución normal. Por lo tanto, es razonable esperar, a pesar de que las dos expresiones sean iguales, que un valor de n obtenido mediante el empleo de (8.36) será mucho más pequeño que el correspondiente valor que se obtiene mediante el empleo de (8.6), debido a que para (8.36) se

TABLA 8.1 Intervalos de confianza para el ejemplo 8.9

Confianza	$z_{1-\alpha/2}$	*Límite inferior*	*Límite superior*
90%	1.645	501.69	505.81
95%	1.96	501.30	506.20
99%	2.575	500.53	506.97

formularon más hipótesis. Para comparar, si se supone que se está muestreando una distribución normal, el tamaño de la muestra que corresponde a las condiciones dadas en el ejemplo 8.2, podría ser

$$n = \frac{(1.645)^2\, 10}{4} \simeq 7,$$

comparado con el valor de $n = 25$ dado por (8.6).

Desde el punto de vista de la aplicación, el hecho de que ambas expresiones tengan como hipótesis el conocimiento de la varianza de la población σ^2, constituye un requisito muy severo. Si no se conoce el valor de σ^2 debe usarse un estimado de σ^2 que quizá pueda encontrarse en una muestra previa. Si este estimado no se encuentra disponible pero se conoce, en forma aproximada, el intervalo en el cual se encuentran las mediciones, una estimación muy burda de la desviación estándar es igual a la sexta parte del recorrido de las observaciones, ya que para muchas distribuciones unimodales la gran mayoría de las observaciones se encontrarán dentro de un intervalo igual a tres desviaciones estándar, ya sea a la izquierda o la derecha de la media.

8.4.2 Intervalos de confianza para μ cuando se muestrea una distribución normal con varianza desconocida

Se considerará el problema de encontrar un intervalo de confianza para μ, cuando se muestrea una distribución normal y para la cual no se tiene conocimiento acerca del valor de la varianza. De la sección 7.6, recuérdese que cuando se muestrea una $N(\mu, \sigma)$, en donde tanto μ como σ^2 son desconocidos, la variable aleatoria

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \tag{8.37}$$

tiene una distribución t de Student con $n - 1$ grados de libertad. Por lo tanto, es posible determinar el valor cuantil $t_{1-\alpha/2,\, n-1}$ de T, para el cual

$$P(-t_{1-\alpha/2,\, n-1} < T < t_{1-\alpha/2,\, n-1}) = 1 - \alpha, \tag{8.38}$$

en donde el valor cuantil es tal que $P(T < -t_{1-\alpha/2,\, n-1}) = \alpha/2$ y $P(T < t_{1-\alpha/2,\, n-1}) = 1 - \alpha/2$. Al sustituir para T en (8.38), se tiene

$$P\left(-t_{1-\alpha/2,\, n-1} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{1-\alpha/2,\, n-1}\right) = 1 - \alpha$$

o

$$P\left(-t_{1-\alpha/2,\, n-1}\frac{S}{\sqrt{n}} < \bar{X} - \mu < t_{1-\alpha/2,\, n-1}\frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

y

$$P\left(\bar{X} - t_{1-\alpha/2,\, n-1}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{1-\alpha/2,\, n-1}\frac{S}{\sqrt{n}}\right) = 1 - \alpha. \tag{8.39}$$

Por lo tanto, el intervalo $\overline{X} \pm t_{1-\alpha/2,\, n-1}(S/\sqrt{n})$ es un intervalo aleatorio y la probabilidad de que éste contenga el valor verdadero de μ, es $1 - \alpha$. De esta forma, dados los datos de una muestra aleatoria de tamaño n a partir de los cuales se calculan los estimados $\bar{x}$ y s^2, un intervalo de confianza del $100(1 - \alpha)\%$ para μ es

$$\bar{x} \pm t_{1-\alpha/2,\, n-1} \frac{s}{\sqrt{n}}. \tag{8.40}$$

Con propósitos de ilustración y comparación, la tabla 8.2 lista los intervalos de confianza del 90, 95 y 99% para μ, con base en (8.40) y mediante el empleo de los datos del ejemplo 8.9, en donde $\bar{x} = 503.75$ y $s = 6.20$. Nótese que para el caso que involucra a la distribución t de Student, los intervalos son más amplios.

8.4.3 Intervalos de confianza para la diferencia de medias cuando se muestrean dos distribuciones normales independientes

Sean $X_1, X_2, \ldots, X_{n_X}$ y $Y_1, Y_2, \ldots, Y_{n_Y}$ dos muestras aleatorias de dos distribuciones normales independientes, con medias μ_X y μ_Y y varianzas σ_X^2 y σ_Y^2, respectivamente. Se desea construir un intervalo de confianza para la diferencia $\mu_X - \mu_Y$. Supóngase que se conocen los valores de las varianzas. Entonces, de la sección 7.7, la variable aleatoria

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}} \tag{8.41}$$

es $N(0, 1)$. De esta forma es posible encontrar el valor cuantil $z_{1-\alpha/2}$, tal que

$$P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = 1 - \alpha. \tag{8.42}$$

Mediante la sustitución de (8.41) en (8.42) y después de manipular algebraicamente las desigualdades, se tiene

$$P\left(\overline{X} - \overline{Y} - z_{1-\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} < \mu_X - \mu_Y < \overline{X} - \overline{Y} + z_{1-\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}\right) = 1 - \alpha, \tag{8.43}$$

TABLA 8.2 Intervalos de confianza para el ejemplo 8.9

Confianza	$t_{1-\alpha/2,\, n-1}$	*Límite inferior*	*Límite superior*
90%	1.753	501.03	506.47
95%	2.131	500.45	507.05
99%	2.947	499.18	508.32

que es un intervalo aleatorio que no contiene parámetros desconocidos. Al igual que en el caso de la sección 8.4.1, la variable aleatoria pivotal es la normal estándar Z. De acuerdo con lo anterior, un intervalo de confianza del $100(1 - \alpha)\%$ para $\mu_X - \mu_Y$ es

$$\bar{x} - \bar{y} \pm z_{1-\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}, \tag{8.44}$$

en donde el valor cuantil $z_{1-\alpha/2}$, es tal que $P(Z < z_{1-\alpha/2}) = 1 - \alpha/2$.

Si las varianzas σ_X^2 y σ_Y^2 se desconocen pero son iguales, entonces la variable aleatoria

$$T = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_p\sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}}$$

tiene una distribución t de Student con $k = n_X + n_Y - 2$ grados de libertad. Al seguir el procedimiento anterior, se tiene que un intervalo de confianza del $100(1 - \alpha)\%$ para $\mu_X - \mu_Y$, es

$$\bar{x} - \bar{y} \pm t_{1-\alpha/2,k}\, s_p\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}, \tag{8.45}$$

en donde el estimado combinado de la varianza común es

$$s_p^2 = \frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2}.$$

Ejemplo 8.10 Se piensa que los estudiantes de licenciatura de contaduría pueden esperar un mayor salario promedio al egresar de la licenciatura, que el que esperan los estudiantes de administración. Recientemente se obtuvieron muestras aleatorias de ambos grupos de un área geográfica relativamente homogénea, proporcionando los datos que se encuentran en la tabla 8.3. Determinar un intervalo de confianza unilateral inferior del 90% para la diferencia entre los salarios promedio para los estudiantes de contaduría y los de administración $\mu_A - \mu_M$ al egresar de la licenciatura (suponga que las varianzas σ_A^2 y σ_M^2 son iguales).

A partir de los datos muestrales dados, pueden calcularse las siguientes cantidades:

$$n_A = 10 \qquad n_M = 14$$

$$\bar{x}_A = 16\,250 \qquad \bar{x}_M = 15\,400$$

$$s_A^2 = 1\,187\,222.22 \qquad s_M^2 = 1\,352\,307.69$$

$$s_p^2 = 1\,284\,772.73$$

$$s_p = 1133.48.$$

TABLA 8.3 Salarios anuales iniciales para recién graduados

Contadores	*Administradores*
$16 300	$13 200
18 200	15 100
17 500	13 900
16 100	14 700
15 900	15 600
15 400	15 800
15 800	14 900
17 300	18 100
14 900	15 600
15 100	15 300
	16 200
	15 200
	15 400
	16 600

Entonces, un intervalo de confianza unilateral inferior del 90% está dado por

$$\bar{x}_A - \bar{x}_M - t_{0.9,\ 22}\, s_p\sqrt{\frac{1}{n_A} + \frac{1}{n_M}},$$

en donde el valor $t_{0.9,\ 22} = 1.321$, ya que para la distribución t de Student, $P(T < 1.321) = 0.9$. Al Sustituir los resultados numéricos, se tiene

$$16\,250 - 15\,400 - (1.321)(1133.48)\sqrt{\frac{1}{10} + \frac{1}{14}} = 230.05.$$

De esta forma, un intervalo de confianza unilateral del 90% para la diferencia real entre los salarios promedio es de $230.05.

8.4.4 Intervalos de confianza para σ^2 cuando se muestrea una distribución normal con media desconocida

Se examinará el problema de construcción de un intervalo de confianza para la varianza de la población σ^2 cuando se muestrea $N(\mu, \sigma)$. De la sección 7.5, se recordará que bajo estas condiciones, la distribución de muestreo de $(n-1)S^2/\sigma^2$ es chi-cuadrada con $n - 1$ grados de libertad. Entonces es posible determinar los valores cuantiles $\chi^2_{\alpha/2,\ n-1}$ y $\chi^2_{1-\alpha/2,\ n-1}$, tales que

$$P\left[\chi^2_{\alpha/2,\ n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{1-\alpha/2,\ n-1}\right] = 1 - \alpha. \qquad (8.46)$$

Puede expresarse (8.46) como

$$P\left[\frac{1}{\chi^2_{\alpha/2,\ n-1}} > \frac{\sigma^2}{(n-1)S^2} > \frac{1}{\chi^2_{1-\alpha/2,\ n-1}}\right] = 1 - \alpha.$$

Entonces el intervalo

$$\left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,\,n-1}}, \frac{(n-1)S^2}{\chi^2_{\alpha/2,\,n-1}}\right]$$

es un intervalo aleatorio el cual contiene a σ^2 y a parámetros conocidos con una probabilidad de $1 - \alpha$. De esta forma, con base en los datos de una muestra aleatoria de tamaño n, se calcula el estimado s^2 y un intervalo de confianza del $100(1 - \alpha)\%$ para σ^2, es de $(n-1)s^2/\chi^2_{1-\alpha/2,\,n-1}$ a $(n-1)s^2/\chi^2_{\alpha/2,\,n-1}$. Es interesante notar que la variable aleatoria pivotal es $(n-1)S^2/\sigma^2$ ya que su función de densidad, dada por (7.16), no contiene ningún parámetro desconocido.

Ejemplo 8.11 Un proceso produce cierta clase de cojinetes de bola cuyo diámetro interior es de 3 cm. Se seleccionan, en forma aleatoria, 12 de estos cojinetes y se miden sus diámetros internos, que resultan ser 3.01, 3.05, 2.99, 2.99, 3.00, 3.02, 2.98, 2.99, 2.97, 2.97, 3.02 y 3.01. Suponiendo que el diámetro es una variable aleatoria normalmente distribuida, determinar un intervalo de confianza del 99% para la varianza σ^2.

Dado que la confianza deseada es del 99%, $\alpha = 0.01$. De la tabla E del apéndice, los valores cuantiles $\chi^2_{0.005,\,11}$ y $\chi^2_{0.995,\,11}$ son 2.60 y 26.71, respectivamente. Para terminar, el valor calculado de la varianza muestral es $s^2 = 0.0005455$. Por lo tanto, un intervalo de confianza del 99% para σ^2 es

$$\left[\frac{(12-1)(0.0005455)}{26.71}, \frac{(12-1)(0.0005455)}{2.60}\right],$$

o

$$(0.0002246, 0.0023079).$$

Como lo ilustra este ejemplo, el punto medio de un intervalo de confianza para una varianza no coincide con el valor del estimador puntual. Sin embargo, cuando se construye un intervalo simétrico como lo es el de la media cuando se muestrea una distribución normal, el punto medio del intervalo de confianza coincide con el estimador puntual.

8.4.5 Intervalos de confianza para el cociente de dos varianzas cuando se muestrean dos distribuciones normales independientes

En el medio industrial muchas veces surge la necesidad de medir y comparar las variabilidades de dos procesos distintos. Supóngase que se tienen muestras aleatorias provenientes de dos distribuciones normales con medias y varianzas desconocidas. Sean n_X y n_Y, el tamaño de las muestras y S_X^2 y S_Y^2 las varianzas muestrales. El interés se centra en construir un intervalo de confianza para el cociente σ_Y^2/σ_X^2 de las dos varianzas poblacionales. De la sección 7.8, se recordará que la variable aleatoria $(S_X^2/\sigma_X^2)/(S_Y^2/\sigma_Y^2)$ tiene una distribución F con $n_X - 1$ y $n_Y - 1$ grados de libertad. Entonces puede escribirse

$$P\left(a < \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} < b\right) = 1 - \alpha, \tag{8.47}$$

en donde a y b son los valores cuantiles inferior y superior de una distribución F tales que

$$a = 1/f_{1-\alpha/2,\, n_Y-1,\, n_X-1} \quad \text{y} \quad b = f_{1-\alpha/2,\, n_X-1,\, n_Y-1}.$$

La proposición de probabilidad dada por (8.47) se puede expresar como

$$P\left(a < \frac{S_X^2}{S_Y^2} \cdot \frac{\sigma_Y^2}{\sigma_X^2} < b\right) = 1 - \alpha$$

o

$$P\left(\frac{aS_Y^2}{S_X^2} < \frac{\sigma_Y^2}{\sigma_X^2} < \frac{bS_Y^2}{S_X^2}\right) = 1 - \alpha. \tag{8.48}$$

De esta manera, un intervalo de confianza del $100(1 - \alpha)\%$ para σ_Y^2/σ_X^2 está dado por

$$(as_Y^2/s_X^2,\; bs_Y^2/s_X^2).$$

Para ilustrar, recuérdese el ejemplo 8.10. Supóngase que se desea un intervalo de confianza del 90% para σ_M^2/σ_A^2. De la tabla G, los valores cuantiles son

$$a = 1/f_{0.95,\, 13,\, 9} = 1/3.05^* = 0.328,$$

$$b = f_{0.95,\, 9,\, 13} = 2.71.$$

Ya que $s_A^2 = 1\ 187\ 222.22$ y $s_M^2 = 1\ 352\ 307.69$, un intervalo de confianza del 90% para el cociente σ_M^2/σ_A^2 de las dos varianzas desconocidas es

$$[(0.328)(1\ 352\ 307.69)/1\ 187\ 222.22,\quad (2.71)(1\ 352\ 307.69)/1\ 187\ 222.22]$$

o

$$(0.3736,\ 3.0868).$$

8.4.6 Intervalos de confianza para el parámetro de proporción p cuando se muestrea una distribución binomial

El porcentaje de productos defectuosos de un proceso de manufactura es el barómetro más importante para medir la calidad del proceso para manufacturar un producto dado. Ya que un artículo puede estar defectuoso o no, el número de unidades defectuosas es una variable aleatoria binomial, si se supone una probabilidad constante e independencia. En una muestra aleatoria de tamaño n el parámetro p que representa la proporción de artículos defectuosos es desconocido. Se desea determi-

* Por interpolación.

nar un intervalo de confianza para p. A pesar de que es posible determinar intervalos de confianza exactos para p (véase [2]), se optará por un intervalo de confianza basado en una muestra grande. La razón de esta decisión tiene sus raíces en el teorema 5.1, el cual establece que la distribución de una variable aleatoria binomial tiende hacia una normal cuando n tiende a infinito.

Se demostró en el ejemplo 8.6 que el estimador de máxima verosimilitud de p, denotado por $\hat{P}$, es

$$\hat{P} = X/n, \tag{8.49}$$

en donde X es binomial con parámetros n y p. Nótese que $\hat{P}$ es un estimador insesgado de p, ya que

$$E(\hat{P}) = \frac{1}{n}E(X) = np/n = p.$$

La varianza de $\hat{P}$ se puede obtener de la siguiente forma:

$$\begin{aligned} Var(\hat{P}) &= Var(X/n) \\ &= \frac{1}{n^2}[np(1 - p)] \\ &= p(1 - p)/n. \end{aligned} \tag{8.50}$$

Recuérdese que para n grande, la variable aleatoria $(X - np)/\sqrt{np(1 - p)}$ es aproximadamente $N(0, 1)$. Entonces puede demostrarse que la distribución de

$$\frac{\hat{P} - p}{\sqrt{\dfrac{\hat{P}(1 - \hat{P})}{n}}} \tag{8.51}$$

también tiende a $N(0, 1)$ para n grande. De esta forma, la probabilidad del intervalo aleatorio

$$\left[\hat{P} - z_{1-\alpha/2}\sqrt{\frac{\hat{P}(1 - \hat{P})}{n}},\quad \hat{P} + z_{1-\alpha/2}\sqrt{\frac{\hat{P}(1 - \hat{P})}{n}}\right] \tag{8.52}$$

es, en forma aproximada, $1 - \alpha$ para n grande. De acuerdo con lo anterior, un intervalo de confianza aproximado del 100 $(1 - \alpha)\%$ para el parámetro de proporción p, es

$$\left[\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}},\quad \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}\right], \tag{8.53}$$

en donde el estimador de máxima verosimilitud $\hat{p} = x/n$ se obtiene de la muestra aleatoria de tamaño n.

Ejemplo 8.12 Un fabricante asegura, a una compañía que le compra un producto en forma regular, que el porcentaje de productos defectuosos no es mayor del 5%. La compañía decide comprobar la afirmación del fabricante seleccionando, de su inventario, 200 unidades de este producto y probándolas. ¿Deberá sospechar la compañía de la afirmación del fabricante si se descubren un total de 19 unidades defectuosas en la muestra?

La sospecha estará apoyada si existe un intervalo de confiabilidad alta para el cual la proporción p se encuentra completamente a la derecha del valor asegurado 0.05. Se selecciona una confiabilidad del 95%. Dado que la realizacion de la variable aleatoria X es $x = 19$ y $n = 20$, el estimado de p es $19/200 = 0.095$. Al sustituir en (8.53), se tiene

$$\left[0.095 - 1.96\sqrt{\frac{0.095(1-0.095)}{200}},\quad 0.095 + 1.96\sqrt{\frac{0.095(1-0.095)}{200}}\right],$$

el cual resulta ser (0.5436, 0.1356). Aparentemente existe una razón para sospechar de la afirmación del fabricante, ya que el intervalo de confianza se encuentra completamente a la derecha del valor asegurado.

Con respecto al muestreo de una distribución binomial, un problema que surge, en forma frecuente, es el de estimar el tamaño de la muestra necesario de manera tal que con una confiabilidad de $100(1 - \alpha)\%$ aproximadamente, el estimado del parámetro de proporción se encuentre a no más de ε unidades de p. Dado el estimador de máxima verosimilitud X/n y siguiendo el mismo procedimiento de la sección 8.4.1, puede expresarse (8.52) como

$$P\left(\left|\frac{X}{n} - p\right| < \varepsilon\right) \simeq 1 - \alpha,$$

en donde

$$\varepsilon = z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

Al resolver para n, se obtiene

$$n = [z^2_{1-\alpha/2}\, p(1 - p)]/\varepsilon^2. \tag{8.54}$$

Nótese que en la expresión anterior n es una función del valor deseado de p. Dado que éste no se conoce y, de hecho, es la cantidad que se está intentando estimar, lo que de manera general se hace es determinar el valor más conservador de n. Esto ocurre cuando la cantidad $p(1 - p)$ es máxima. Pero puede demostrarse que para $0 \leq p \leq 1$, $p(1 - p)$ es un máximo cuando $p = 1/2$. En otras palabras, el valor $p = 1/2$ es el que debe emplearse para obtener el tamaño deseado de la muestra con base en (8.54).

A manera de advertencia, los métodos presentados en esta sección deben usarse sólo cuando el tamaño de la muestra es suficientemente grande. De otro modo, de-

berán emplearse los intervalos de confianza exactos. A lo largo de estos lineamientos, de nuevo debe hacerse énfasis en que se supuso que el muestreo siempre se llevaba a cabo sobre una distribución normal. La construcción de intervalos de confianza para las varianzas es, en forma especial, sensible a esta hipótesis. Cualquier desviación sustantiva de esta hipótesis significará una pérdida de la validez de la inferencia formulada con respecto a las varianzas. Por otro lado, los métodos para intervalos de confianza que involucran medias son, en forma relativa, inmunes a modestas desviaciones de la hipótesis de normalidad siempre y cuando el tamaño de las muestras sea grande. De esta forma, los métodos presentados de la sección 8.4.1 a la 8.4.3, tienen una gran validez para muestras de gran tamaño, aun si el muestreo no se lleva a cabo sobre una distribución normal.

Para ilustrar que el uso de la distribución t de Student sigue siendo válido para inferencias con respecto a las medias, aun a pesar de que se muestree una distribución que no es normal, se simuló el siguiente experimento mediante el empleo del paquete IMSL. Se generaron 1 000 muestras de tamaños 15, 30 y 50, a partir de una distribución exponencial con parámetro $\theta = 10$. Ya que θ es la media de una variable aleatoria exponencialmente distribuida, se empleó (8.40) para calcular un intervalo de confianza del 95% para θ para cada muestra aleatoria y se contó el número de intervalos que no contenían el valor supuesto de 10. Para $n = 15$ se encontró un total de 86 de estos intervalos; para $n = 30$ se tienen 68 y para $n = 50$ se encontraron 55.

Si el muestreo se hubiese llevado a cabo sobre una distribución normal, se esperarían (0.05) (1 000) = 50 de estos intervalos, de entre 1 000. Parece ser que los resultados se acercan a los esperados bajo un muestreo de una distribución normal conforme aumenta el tamaño de la muestra aun a pesar de que ésta no provenga de una distribución normal. De acuerdo con lo anterior, el efecto que se tiene por una violación de la hipótesis de normalidad cuando se utiliza la distribución t de Student, parece ser pequeño, aun para un tamaño n relativamente modesto.

8.5 Estimación bayesiana

Hasta este momento se ha estudiado la inferencia estadística desde el punto de vista de la teoría del muestreo, el cual se basa en la interpretación de la probabilidad como una frecuencia relativa. En esta sección se estudiará el enfoque bayesiano de la inferencia estadística y, en particular, a la estimación de parámetros. Recuérdese que el enfoque bayesiano se basa en la interpretación subjetiva de la probabilidad, el cual considera a ésta como un grado de creencia con respecto a la incertidumbre. El punto de vista bayesiano considera un parámetro desconocido como una característica con respecto a la cual puede expresarse un grado de creencia que puede modificarse con base en la informacion muestral. Una inferencia con respecto al parámetro se formula con base en el grado de creencia modificado. En otras palabras, un parámetro es visto como una variable aleatoria a la que, antes de la evidencia muestral, se le asigna una distribución *a priori* con base en el grado de creencia con respecto al comportamiento del parámetro aleatorio. Cuando se obtiene la evidencia muestral, la distribución *a priori* es modificada y entonces surge una distribución *a posteriori*. Es esta distribución *a posteriori* la que se emplea para formular inferencias con respecto al parámetro.

El enfoque bayesiano para la estimación de parámetros ha sido favorecido por muchas personas, en forma especial en aquellas situaciones en las que un parámetro no puede considerarse, en forma real, como una cantidad fija. Por ejemplo, es probable que la verdadera proporción de artículos defectuosos que produce un proceso de manufactura fluctúe ligeramente, lo cual depende de numerosos factores, como se mostró en el ejemplo 6.9. Es probable que la verdadera proporción de casas que se pierden por concepto de hipoteca varíe dependiendo, en primer lugar, de las condiciones económicas. La demanda promedio semanal de automóviles también fluctuará como una función de varios factores incluyendo la temporada.

8.5.1 Estimación puntual bayesiana

En esta sección se considerará la determinación de estimadores puntuales bayesianos. Dado que se considera a un parámetro como una variable aleatoria, se denotará a éste por el símbolo Θ y con θ a la realización de Θ. Supóngase que Θ es una variable aleatoria continua* con una función de densidad (*a priori*) incondicional $f_\Theta(\theta)$, la cual refleja la creencia *a priori* con respecto a la incertidumbre de Θ. La información muestral se encuentra representada por n variables aleatorias IID $X_1, X_2, \ldots, X_n$, con una densidad $f(x \mid \theta)$ condicional común sobre la realización θ de Θ. Del capítulo 7, la función de verosimilitud, condicional a un valor particular θ, es

$$L(x_1, x_2, \ldots, x_n \mid \theta) = f(x_1 \mid \theta) f(x_2 \mid \theta) \cdots f(x_n \mid \theta). \tag{8.55}$$

Es importante hacer énfasis en que aun cuando Θ es una variable aleatoria, el objetivo es estimar el valor particular de θ para el cual la evidencia muestral que representa la función de verosimilitud se encuentra condicionada. Es decir, Θ es una variable aleatoria no observable que puede tomar varios valores (entre ellos θ,) que deriven el resultado muestral. Mediante el empleo del teorema 6.2 y, en particular, de (6.24), la densidad *a posteriori* de Θ dado el resultado muestral $\underline{x} = \{x_1, x_2, \ldots, x_n\}$ es

$$f(\theta \mid \underline{x}) = \frac{L(\underline{x} \mid \theta) f_\Theta(\theta)}{\int_\Theta L(\underline{x} \mid \theta) f_\Theta(\theta) d\theta}. \tag{8.56}$$

Se sabe que la densidad *a posteriori* $f(\theta \mid \underline{x})$ representa el grado de creencia modificado con respecto a la incertidumbre de Θ. Pero ¿cómo debe usarse la densidad *a posteriori* para obtener un estimador puntual de θ? Para este propósito, el enfoque bayesiano** toma en cuenta una *función de pérdida*, que representa la consecuencia económica resultante de haber escogido a $t = u(\underline{x})$ como el valor estimado cuando el valor verdadero es θ. Esto es, la función de pérdida evalúa la pérdida económica cuando se dice que el valor de θ es t, cuando éste es θ. Una función de pérdida, denotada por $l(t, \theta)$, es una función no negativa de t y θ de tal forma que ésta es cero

* Es más probable que un parámetro desconocido sea continuo que discreto, pero este último caso puede manejarse en forma similar.

** Para una presentación más completa del enfoque bayesiano se invita al lector a que consulte [6].

sólo si t es igual a θ. Nótese que la función de pérdida depende del parámetro aleatorio Θ; por lo tanto, ésta también es una variable aleatoria. En este momento se está en condiciones de definir un estimador bayesiano.

Definición 8.8 Sea $f_\Theta(\theta)$ la función de densidad *a priori* de un parámetro Θ, y $L(x_1, x_2, \ldots, x_n \mid \theta)$ la función de máxima verosimilitud de una muestra aleatoria de n variables aleatorias IID condicionadas sobre la realización θ de Θ. Además, sea $f(\theta \mid \underline{x})$ la función de densidad *a posteriori* de Θ, y sea $l(t, \Theta)$ la función de pérdida. El estimador Bayes de θ, $T = u(X_1, X_2, \ldots, X_n)$, es aquél para el cual el valor esperado de la función de pérdida dada por

$$E[l(t, \Theta)] = \int_\Theta l(t, \theta) f(\theta \mid \underline{x}) d\theta$$

es mínimo.

En la definición 8.8 es claro que para determinar un estimador Bayes, debe especificarse una función de pérdida. La especificación de esta última es una tarea difícil, ya que las consecuencias económicas no son fácilmente medibles. En muchos problemas de aplicación puede formularse un argumento razonable para utilizar una función de pérdida de la forma.

$$l(t, \theta) = (t - \theta)^2, \tag{8.57}$$

la cual se conoce como función de pérdida *cuadrática* o de *error cuadrático*. Para una función de pérdida cuadrática puede demostrarse que el estimador Bayes de θ es igual a la esperanza *a posteriori* $E(\Theta \mid \underline{x})$, de Θ. En otras palabras, la media de la distribución *a posteriori* de Θ es el estimador Bayes de θ para una función de pérdida de error cuadrático. Nótese que ésta es una elección razonable para estimar el valor de la realización θ, ya que la media de una variable aleatoria es una medida de tendencia central y representa el centro de gravedad de la distribución de probabilidad de la variable aleatoria.

Ejemplo 8.13 Un vendedor distribuye sistemas estereofónicos, los cuales garantiza por un periodo de dos años. Con base en información previa, el vendedor piensa que la proporción de unidades que serán enviadas a servicio o a reemplazo durante el periodo de dos años tiene un valor cercano a 0.04, aunque existen ligeras variaciones de este valor. El vendedor piensa asignar *a priori* una distribución beta a la proporción con parámetros $\alpha = 1$ y $\beta = 24$. Con base en una muestra aleatoria de 25 unidades, el vendedor observa dos unidades que necesitarán servicio o reemplazo durante el periodo de dos años. Suponiendo que el número de unidades que necesitarán, ya sea servicio o reemplazo en una muestra fija de n unidades, es una variable aleatoria binomial, obtener el estimador Bayes de la proporción.

En el ejemplo 6.9, se demostró que, para las condiciones de este problema, la distribución *a posteriori* de la proporción también es una distribución beta con una densidad dada por (6.36). Denótese a la proporción aleatoria por P. Ya que los pará-

metros de la densidad *a posteriori* de P son $x + \alpha$ y $n + \beta - x$, y mediante el empleo de (5.40), la media *a posteriori*.

$$E(P \mid x) = \frac{x + \alpha}{n + \alpha + \beta} \tag{8.58}$$

es el estimador Bayes de la realización p. Antes de calcular el valor del estimador, es conveniente comparar el estimador Bayes con el estimador de máxima verosimilitud x/n, que se obtuvo en el ejemplo 8.6. Nótese que el estimador Bayes coincide con el de máxima verosimilitud sólo si $\alpha = \beta = 0$. Para este problema el resultado muestral para $n = 25$ es $x = 2$, y los valores de los parámetros *a priori* son $\alpha = 1$ y $\beta = 24$. De esta forma, el estimador Bayes es $(2 + 1)/(25 + 1 + 24) = 0.06$, y por comparación, el estimador MV es $2/25 = 0.08$.

Por lo tanto, es evidente que el estimador Bayes se encuentra influenciado tanto por el resultado muestral como por la distribución *a priori*. De hecho, puede decirse que si la distribución *a priori* tiene una varianza pequeña, lo que implica un alto grado de creencia con respecto a un parámetro aleatorio, entonces la media *a posteriori* tendrá un valor muy próximo a la media *a priori*. Supóngase, para el ejemplo 8.13, que los valores de α y β fuesen 2 y 48 en lugar de 1 y 24, respectivamente. Entonces el valor de la media *a priori* debería ser igual al que se dio en $2/(2 + 48) = 0.04$ pero la varianza *a priori* debe ser, ahora, igual a 0.0007529, que es un valor más pequeño que el anterior (0.0014769). El resultado es la media $(2 + 2)/(25 + 2 + 48) = 0.0533$ y se encuentra más cercano al valor de la media *a priori* que el estimado previo. Por otro lado, si la distribución *a priori* tiene una varianza muy grande, ésta debe ser virtualmente plana, lo cual implica que la creencia *a priori* con respecto a la incertidumbre de un parámetro aleatorio es vaga. En tal caso, la evidencia muestral debe tener mucho más peso en la distribución *a posteriori* que en la distribución *a priori*, y los estimadores de Bayes y MV deberán ser, virtualmente, los mismos.

El tamaño de la muestra n también tiene influencia sobre la cercanía entre los estimadores Bayes y MV. En general, los estimadores Bayes y MV diferirán entre sí por una cantidad que es pequeña cuando se compara con $1/\sqrt{n}$. De esta manera, para tamaños de la muestra relativamente grandes ambos estimadores se encontrarán muy cercanos el uno del otro.

8.5.2 Estimación bayesiana por intervalo

Se puede determinar un intervalo estimado para θ mediante el uso de la función de densidad *a posteriori* del parámetro aleatorio Θ.

Definición 8.9 Sea $f(\theta \mid \underline{x})$ la función de densidad *a posteriori* de Θ condicionada sobre el resultado muestral t $\underline{x} = \{x_1, x_2, \ldots, x_n\}$, sean a y b límites tales que

$$P(a < \Theta < b \mid \underline{x}) = \int_a^b f(\theta \mid \underline{x}) d\theta = \gamma, \tag{8.59}$$

en donde a y b son funciones del resultado muestral $\underline{x}$. Entonces el intervalo (a, b) es un intervalo bayesiano tal, que la probabilidad de que θ se encuentre contenido en (a, b) es γ.

A diferencia de los intervalos de confianza de la sección 8.4, un intervalo bayesiano es, en efecto, un intervalo de probabilidad. En otras palabras, puede decirse que la probabilidad de que γ se encuentre contenido en el intervalo a, b es θ, mientras que con un intervalo de confianza sólo puede decirse que una cantidad de $100\gamma\%$ de estos intervalos contendrán el valor real de θ.

Para ejemplificar un intervalo de probabilidad bayesiano, sea $X_1, X_2, \ldots, X_n$ la muestra aleatoria de una distribución normal con media μ desconocida y varianza σ^2 conocida. Supóngase que la media es un parámetro aleatorio al cual se piensa asignar una distribución normal *a priori* con una función de densidad

$$f_M(\mu) = \frac{1}{\sigma_0\sqrt{2\pi}}\exp[-(\mu - \mu_0)^2/2\sigma_0^2] \qquad -\infty < \mu < \infty,$$

donde μ_0 y σ_0^2 son la media y la varianza *a priori*, respectivamente. De la presentación previa (véase el ejemplo 8.7), la función de verosimilitud dada la realización μ es

$$L(x_1, x_2, \ldots, x_n \mid \mu) = (2\pi\sigma^2)^{-n/2}\exp[-\Sigma(x_i - \mu)^2/2\sigma^2].$$

Entonces, puede demostrarse que la densidad *a posteriori* de la media condicionada sobre $\underline{x}$ también es normal con media

$$E(M \mid \underline{x}) = \frac{n\sigma_0^2\bar{x} + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2} \tag{8.60}$$

y varianza

$$Var(M \mid \underline{x}) = \frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2}. \tag{8.61}$$

De esta forma, el estimador Bayes de μ para una función de pérdida o error cuadrático está dada por (8.60). Al igual que en el ejemplo 8.13, nótese que un valor pequeño de la varianza *a priori* σ_0^2 proporcionará un estimador Bayes para μ mucho más cercano a la media *a priori* μ_0. Además, para μ_0 y σ_0^2, fijas, conforme n crece el estimador de Bayes tiende al estimador de máxima verosimilitud $\bar{x}$.

Ejemplo 8.14 Recuérdese el ejemplo 8.9 en el que se determinaron los intervalos de confianza del 90, 95 y 99% para el llenado medio μ con base en los pesos de 16 cajas de cereal seleccionadas en forma aleatoria y en donde se supuso que los pesos estaban normalmente distribuidos con $\sigma = 5$ gr. Debido a pequeñas perturbaciones en el proceso de llenado, supóngase que el llenado medio es una variable aleatoria normalmente distribuida con media $\mu_0 = 500$ y desviación estándar $\sigma_0 = 1$. Determinar los intervalos de probabilidad bayesiana 0.9, 0.95 y 0.99 para μ.

Del ejemplo 8.9, $\bar{x} = 503.75$; entonces, mediante el uso de (8.60) y (8.61), los valores calculados de la media y la varianza *a posteriori* son

$$E(M \mid \underline{x}) = \frac{(16)(1)(503.75) + (500)(25)}{(16)(1) + 25} = 501.4634$$

y

$$Var(M \mid \underline{x}) = \frac{(25)(1)}{(16)(1) + 25} = 0.6098,$$

respectivamente. Dado que la densidad *a posteriori* de M es $N(501.4634, \sqrt{0.6098})$, y ya que para $\gamma = 0.9$, $P(-1.645 < Z < 1.645) = 0.9$, en donde $Z \sim N(0, 1)$, se sigue de (8.59) que un intervalo de probabilidad 0.9 para μ que sea simétrico con respecto a la media *a posteriori* es

$$E(M \mid \underline{x}) \pm 1.645 \sqrt{Var(M \mid \underline{x})}.$$

De esta forma los límites son $a = E(M \mid \underline{x}) - 1.645\sqrt{Var(M \mid \underline{x})}$ y $b = E(M \mid \underline{x}) + 1.645\sqrt{Var(M \mid \underline{x})}$. Al sustituir los valores para $E(M \mid \underline{x})$ y $\sqrt{Var(M \mid \underline{x})}$, se obtiene el intervalo de probabilidad 0.9 (500.18, 502.75) para μ. De manera similar, se calculan los intervalos bayesianos para $\gamma = 0.95$ y $\gamma = 0.99$. Éstos se encuentran resumidos en la tabla 8.4. Nótese que los intervalos de probabilidad bayesianos se estrechan de manera más uniforme que los correspondientes intervalos de confianza del ejemplo 8.9.

8.6 Límites estadísticos de tolerancia

En la sección 5.4 se mencionaron los límites estadísticos de tolerancia y se comentó su importancia para estimar la variabilidad de un producto. En esta sección se desarrollarán límites estadísticos de tolerancia cuando se muestrea una distribución no específica de probabilidad, o cuando el muestreo se lleva a cabo sobre una distribución normal. Estos límites se conocen como límites de tolerancia *independientes de la distribución* debido a que ésta no se especifica.

8.6.1 Límites de tolerancia independientes de la distribución

Imagine un fenómeno aleatorio que involucre la fabricación de un cierto producto. Sea X la variable de medición de este fenómeno, y sea $f(x; \theta)$ la función de densidad de probabilidad de X, en donde θ es un parámetro fijo.

Definición 8.10 Si D es la proporción de observaciones de la variable aleatoria que se encuentra entre los límites L_1 y L_2, que son funciones univaluadas de las observaciones de manera tal que

$$D = \int_{L_1}^{L_2} f(x; \theta)dx = F_X(L_2; \theta) - F_X(L_1; \theta), \tag{8.62}$$

entonces L_1 y L_2 reciben el nombre de *límites estadísticos de tolerancia.*

TABLA 8.4 Intervalos de probabilidad bayesiana para el ejemplo 8.14

Probabilidad	*Límite inferior*	*Límite superior*
0.9	500.18	502.75
0.95	499.93	502.99
0.99	499.45	503.47

Ya que L_1 y L_2 son funciones univaluadas de las observaciones, ellas mismas son variables aleatorias. A su vez, la proporción D es una variable aleatoria, y la proposición de probabilidad

$$P(D \geqslant d) = \gamma$$

tiene un significado que se interpreta como la probabilidad γ de que la proporción de valores en la distribución de X entre L_1 y L_2 no sea menor que d.

Sean $X_{(r)}$ y $X_{(n-r+1)}$ el r-ésimo valor más pequeño y el $(n-r+1)$-ésimo valor más grande, respectivamente, en una muestra aleatoria de tamaño n la cual involucra a la variable de medición X. Se ha demostrado que la proporción de valores D que se encuentran entre $L_1 = X_{(r)}$ y $L_2 = X_{(n-r+1)}$ tiene una distribución beta con parámetros $\alpha = n - 2r + 1$ y $\beta = 2r$, sin importar la forma de la función de densidad de probabilidad de X, en donde L_1 y L_2 son de orden simétrico. De esta forma

$$P(D \geqslant d) = 1 - F_B(d; n - 2r + 1, 2r) = \gamma. \qquad (8.63)$$

La expresión (8.63) es muy fuerte porque permite la determinación de límites estadísticos de tolerancia sin necesidad de especificar la distribución de la variable aleatoria X de interés. Estos límites se conocen como límites de tolerancia independientes de la distribución. Nótese que la relación (8.63) involucra cuatro cantidades, n, r, d y γ. Con el uso de las tablas beta el conocimiento de tres de ellas proporcionará el valor de la cantidad faltante.

El principal uso de (8.63) es determinar el tamaño más pequeño de la muestra de manera tal que con una probabilidad γ por lo menos una proporción d de la distribución de X se encuentre incluida entre los dos valores extremos de la muestra, $X_{(1)}$ y $X_{(n)}$. Esto es, para $r = 1$, (8.63) se reduce a

$$P(D \geqslant d) = 1 - F_B(d; n - 1, 2) = \gamma,$$

la que puede simplificarse para obtener

$$\gamma = 1 - [nd^{n-1} - (n - 1)d^n], \qquad (8.64)$$

lo que da como resultado una expresión en la que puede aparecer la función de distribución beta como una suma si uno de los parámetros de forma es un número entero pequeño (véase [1]).

En la figura 8.2 se dan varias proporciones útiles de d en función del tamaño de la muestra n y la probabilidad γ. Por ejemplo, si se obtiene una muestra de tamaño 25 de una distribución con una función de densidad desconocida, la probabilidad de que por lo menos el 80% de los valores de X se encuentren entre los dos valores extremos de la muestra es de 0.973.

Muchas veces se buscan límites de tolerancia unilaterales de manera tal que la probabilidad de que por lo menos una proporción d de la distribución de X sea más grande de un límite de tolerancia inferior o menor que un límite de tolerancia superior, sea γ. Puede demostrarse, sin importar la distribución de X, que

$$P(D \geqslant d) = 1 - F_B(d; n - r + 1, r) = \gamma. \qquad (8.65)$$

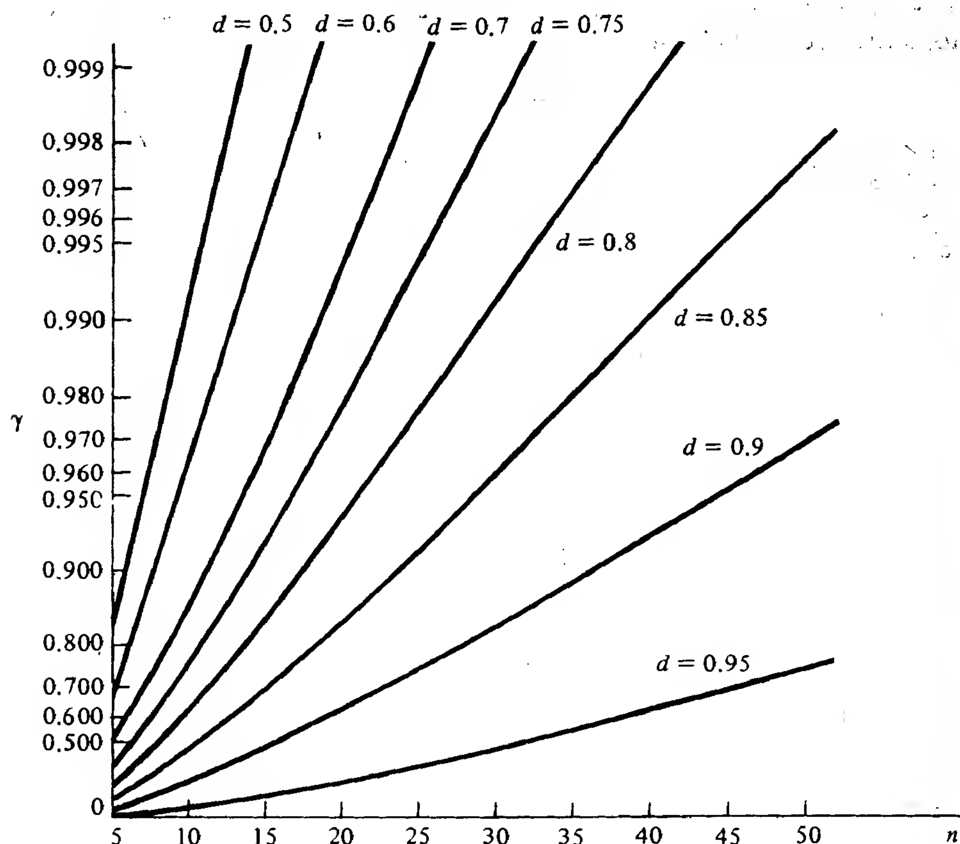


FIGURA 8.2 Proporciones d como funciones del tamaño de muestra n y probabilidad γ.

Nótese que si $r = 1$, la inferencia se formulará con base en el valor más pequeño de la muestra, $X_{(1)}$; si $r = n$, la inferencia se formulará con base en el valor más grande, $X_{(n)}$. Puede demostrarse que, para $r = 1$, la expresión (8.69) se reduce a

$$P(D \geqslant d) = 1 - d^n = \gamma.$$

De esta manera, al obtener el resultado para el tamaño de la muestra n, se tiene

$$n = \frac{\log(1 - \gamma)}{\log(d)}. \tag{8.66}$$

La expresión (8.66) permite la determinación del tamaño necesario de la muestra para que con una probabilidad γ, por lo menos una proporción d de los valores de X sean más grandes que el valor más pequeño de la muestra.

8.6.2 Límites de tolerancia cuando se muestrea una distribución normal

En algunas situaciones la distribución de interés puede modelarse en forma adecuada por una distribución normal. En esta sección se examinarán los límites estadísticos de tolerancia para estas situaciones.

Recuérdese que los límites estadísticos de tolerancia colocan límites sobre las mediciones que se llevan a cabo sobre una distribución a diferencia de los intervalos de confianza, los cuales determinan a aquéllos donde es probable que se encuentre un parámetro desconocido. De esta forma, si el muestreo se lleva a cabo sobre una distribución $N(\mu;\sigma)$ de manera tal que tanto μ como σ son conocidos, entonces, por ejemplo, los límites $\mu \pm 1.645\sigma$, $\mu \pm 1.96\sigma$, y $\mu \pm 2.575\sigma$ incluirán al 90, 95 y 99% de la distribución, respectivamente. O para los límites unilaterales, el 90% de las observaciones de la distribución excederá el límite inferior de $\mu - 1.28\sigma$, y el 99% será menor del límite superior $\mu + 2.33\sigma$. El único problema, con toda seguridad, es que no es muy común el conocer los valores de la media μ y la varianza σ^2.

Supóngase que se consideran los estimadores $\overline{X}$ y S^2. Dado que ambos son variables aleatorias y están sujetas a la variabilidad en el muestreo no es verdad decir, por ejemplo, que el 90% de la distribución estará contenido en el intervalo, $\overline{X} \pm 1.645S$. En forma alternativa, considere el intervalo aleatorio $\overline{X} \pm kS$, en donde k es una constante apropiada perteneciente a la distribución conjunta de $\overline{X}$ y S^2. Dado que $\overline{X} \pm kS$ son límites aleatorios, es imposible establecer con absoluta certeza qué porcentaje de la distribución estará contenido entre estos límites. En otras palabras, al igual que con los intervalos de confianza, no es posible encontrar un valor de k tal que los límites calculados, con base en alguna muestra aleatoria, siempre incluyan un porcentaje fijo de la distribución. Sin embargo, es posible seleccionar un valor de k tal que si se obtienen en forma repetida muestras del mismo tamaño de una distribución normal, proporción fija γ de estos límites contendrá por lo menos un $100d$% de los valores de la distribución. Es decir, el intervalo aleatorio $\overline{X} \pm kS$ tiene una probabilidad γ de contener por lo menos un $100d$% de la distribución normal muestreada. Con base en una muestra aleatoria de tamaño n los límites de tolerancia bilateral de un $100\gamma\%$ para un porcentaje $100d$ de una distribución normal son $\bar{x} \pm ks$, en donde γ es el coeficiente de confianza y d es el alcance de la distribución. La tabla H contiene valores de k para valores seleccionados de n, γ, y d.

Muchas veces sólo se tiene interés en los límites de tolerancia unilaterales. Por ejemplo, en la fabricación de pistones, si el diámetro se encuentra por debajo de cierta tolerancia, el pistón debe desecharse. Sin embargo, si el diámetro del pistón es mayor que cierta tolerancia, éste puede ser reprocesado hasta alcanzar un nivel aceptable. Como era de esperarse, los valores de k para los límites unilaterales no son iguales a los que se encuentran en la tabla H. Éstos se hallan en la tabla I del apéndice para los valores de n, γ, y d más frecuentemente utilizados. De acuerdo con lo anterior, puede determinarse un valor de k tal que, con una confiabilidad del 100γ % de que por lo menos un $100d$% de los valores de la distribución normal serán mayores que el límite de tolerancia inferior $\bar{x} - ks$, o menores que el límite de tolerancia superior $\bar{x} + ks$.

Ejemplo 8.15 **En un medio muy competitivo, la disponibilidad de un producto con respecto a la demanda es crucial para el éxito del negocio. Para determinar un límite de tolerancia superior para la demanda mensual de cierto producto, un centro comercial ha recolectado lo que cree que es una muestra aleatoria de las demandas mensuales y la cual consiste en los siguientes datos: 129, 142, 145, 153, 136, 138, 163, 151, 146, 128, 133, 148, 144, 140, 143. Si la demanda mensual de este producto se encuentra aproximada en forma adecuada por una distribución normal, determínese un límite de tolerancia superior con $\gamma = 0.99$ y $d = 0.95$.**

Para $\gamma = 0.99$, $d = 0.95$ y $n = 15$, se obtiene de la tabla I del apéndice un valor de $k = 3.102$. Con base en los datos, la media y la desviación estándar muestral tienen un valor de $\bar{x} = 142.6$ y $s = 9.2798$, respectivamente. El límite de tolerancia superior es $142.6 + (3.102)(9.2798) = 171.39$. De esta forma, se tiene el 99% de confiabilidad, porque el 95% de toda la demanda será menor que 171.39 unidades por mes. En otras palabras, si este centro comercial almacena aproximadamente 172 unidades del producto por mes, tendrá una alta seguridad de satisfacer la demanda mensual de este producto.

De nuevo, debe hacerse énfasis en que los límites estadísticos de tolerancia desarrollados en esta sección se relacionan con el muestreo de una distribución normal. Si existe alguna duda con respecto a esta hipótesis, deberán utilizarse los límites de tolerancia independientes de la distribución que se estudiaron en la sección 8.6.1. Es razonable esperar que los límites de tolerancia independientes de la distribución sean más conservadores que aquéllos basados en la distribución normal, ya que se encuentra disponible una cantidad menor de información.

Referencias

1. K. V. Bury, *Statistical models in applied science,* Wiley, New York, 1975.
2. R. V. Hogg and A. T. Craig, *Introduction to mathematical statistics,* 4th ed., MacMillan, New York, 1978.
3. A. M. Mood and F. A. Graybill, *Introduction to the theory of statistics,* 2nd ed., McGraw-Hill, New York, 1963.
4. C. R. Rao, *Advanced statistical methods in biometric research,* Wiley, New York, 1952.
5. S. S. Wilks, *Mathematical statistics,* Wiley, New York, 1962.
6. R. L. Winkler, *An introduction to Bayesian inference and decision,* Holt, Rinehart and Winston, New York, 1972.

Ejercicios

8.1. En un experimento binomial se observan x éxitos en n ensayos independientes. Se proponen las siguientes dos estadísticas como estimadores del parámetro de proporción p: $T_1 = X/n$ y $T_2 = (X + 1)/(n + 2)$.

a) Obtener y comparar los errores cuadráticos medios para T_1 y T_2.
b) Hacer una gráfica del ECM de cada estadística como funciones de p para $n = 10$ y $n = 25$. ¿Es alguno de estos estimadores uniformemente mejor que el otro?

8.2. Sea X_1, X_2, X_3, y X_4 una muestra aleatoria de tamaño cuatro de una población cuya distribución es exponencial con parámetro θ desconocido. De las siguientes estadísticas, ¿cuáles son estimadores insesgados de θ?

$$T_1 = \frac{1}{6}(X_1 + X_2) + \frac{1}{3}(X_3 + X_4)$$

$$T_2 = (X_1 + 2X_2 + 3X_3 + 4X_4)/5$$

$$T_3 = (X_1 + X_2 + X_3 + X_4)/4$$

8.3. Demostrar que la estadística T_1, en el ejercicio 8.1, es un estimador consistente del parámetro binomial p.

8.4. Mediante el uso del teorema de Tchebysheff, demostrar que la estadística T_2, en el ejercicio 8.1, es un estimador consistente del parámetro binomial p.

8.5. De entre los estimadores insesgados de θ dados en el ejercicio 8.2, determinar cuál es el que tiene la varianza más pequeña. ¿Cuáles son las eficiencias relativas de los demás estimadores insesgados con respecto al que tiene la varianza más pequeña?

8.6. Sea X_1, X_2, X_3, X_4 y X_5 una muestra aleatoria de una población cuya distribución es normal con media μ y varianza σ^2. Considérense las estadísticas $T_1 = (X_1 + X_2 + \cdots + X_5)/5$ y $T_2 = (X_1 + X_2 + 2X_3 + X_4 + X_5)/6$ como estimadores de μ. Identificar la estadística que posee la varianza más pequeña.

8.7. Mediante el uso de la cota inferior de Cramér-Rao determinar la varianza del estimador insesgado de varianza mínima de θ cuando se muestrea una población cuya distribución es exponencial con una densidad $f(x; \theta) = (1/\theta)\exp(-x/\theta)$, $x > 0$. Deducir que el estimador eficiente de θ es la media muestral.

8.8. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población cuya distribución es gama con parámetro de forma conocido. Demostrar que el estimador de máxima verosimilitud para el parámetro de escala está dado por la expresión (8.8).

8.9. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población cuya distribución es de Poisson con parámetro λ. Obtener el estimador de máxima verosimilitud de λ.

8.10. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población cuya distribución es exponencial con parámetro de escala θ. Obtener el estimador de máxima verosimilitud de θ y demostrar que éste es una estadística suficiente para θ.

8.11. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población cuya distribución es la de Rayleigh, con densidad $f(x; \sigma^2) = (x/\sigma^2)\exp(-x^2/2\sigma^2)$, $x > 0$. Obtener el estimador de máxima verosimilitud de σ^2. ¿Es ésta una estadística para σ^2?

8.12. De manera equivalente a la definición 8.7, se define el r-ésimo momento muestral con respecto a la media, como

$$M_r = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^r}{n},$$

en donde $X_1, X_2, \ldots, X_n$ es una muestra aleatoria. Empléense estos momentos para calcular los factores de forma muestrales para los datos dados en el ejercicio 1.1.

¿Puede formularse cualquier inferencia con respecto a la población de interés con base en los factores de forma?

8.13. Repetir el ejercicio 8.12 usando los datos proporcionados en el ejercicio 1.2.

8.14. La tabla 8.5 es una distribución de frecuencias para accidentes automovilísticos recabada para un estudio en California.* Asumiendo que el número de accidentes es una variable aleatoria binomial negativa, úsese el método de momentos para estimar los parámetros binomiales negativos k y p. Comparar las frecuencias que se observaron con aquéllas que se obtienen mediante el empleo de los valores estimadores de k y de p.

8.15. Los siguientes datos son una muestra aleatoria de duración en horas, que se observaron para un determinado componente eléctrico: 142.84, 97.04, 32.46, 69.14, 85.67, 114.43, 41.76, 163.07, 108.22, 63.28. Supóngase que la duración de un componente es una variable aleatoria de Weibull con parámetro de forma $\alpha = 2$.

a) Obtener un estimador de máxima verosimilitud para el parámetro de escala θ.
b) El método de momentos, ¿daría un estimador de θ diferente al que se obtuvo en la parte *a*?
c) Mediante el uso de su respuesta al inciso *a*, estimar la confiabilidad de este componente para $t = 150$ horas.

8.16. Mediante el uso de su respuesta al inciso *a* del ejercicio 8.15, obtener el tiempo para el cual la confiabilidad del componente es de 0.95.

8.17. Los siguientes datos son tiempos de falla, ordenados en horas de diez componentes que fallarán de un total de 40 en una prueba de duración: 421, 436, 448, 474, 496, 499. 510, 525, 593, 675. Supóngase que el tiempo de falla es una variable aleatoria exponencialmente distribuida.

a) Obtener un estimador de máxima verosimilitud para el parámetro θ.
b) Úsese la respuesta de la parte *a* para estimar la confiabilidad de este componente para $t = 4\,000$ horas.

8.18. Una prueba de duración será terminada cuando fallen $m < n$ unidades. Si el tiempo de falla es una variable aleatoria de Weibull con parámetro de forma conocido, obtener el estimador de máxima verosimilitud para el parámetro de escala θ.

**Multivariate analysis of driver accident frequencies over a period of 14 years*, California Department of Motor Vehicles, FHWA Project No. B0149, 1975.

TABLA 8.5

Número de accidentes	*Número de conductores*
0	35,068
1	13,411
2	4,013
3	1,184
4	353
5	93
6	29
7	8
8	4
9 o más	2

8.19. Se desea obtener un indicador del éxito financiero de ciertas tiendas que venden artículos especiales en los centros comerciales de una gran ciudad. Se selecciona una muestra aleatoria de 30 tiendas ubicadas en distintos centros comerciales y en donde el interés recae en el tiempo que éstas permanecen en operación. Se tendrá un dato significativo cuando se observen las primeras ocho tiendas que dejen de funcionar. Los siguientes datos son el tiempo en orden ascendente, de operación en meses: 3.2, 3.9, 5.9, 6.5, 16.5, 20.3, 40.4, 50.9. Supóngase que el tiempo en el que permanece operando una tienda de esta clase es una variable aleatoria de Weibull con $\alpha = 0.8$.

a) Usando el resultado del ejercicio 8.18, obtener el estimador de máxima verosimilitud para θ.
b) Con base en la respuesta del inciso *a*, ¿cuál es la probabilidad de que una tienda permanezca en operación después de haber transcurrido dos años de su apertura? ¿Después de diez años?

8.20. El tiempo total de procesamiento para programas en tarjetas perforadas de computadora se define como el tiempo que transcurre desde que se lee la primera tarjeta hasta que se imprime la última línea, y está constituido por tres componentes; el tiempo de espera de entrada, el tiempo utilizado por el procesador central y el tiempo de espera de salida. Los siguientes datos son los tiempos totales de procesamiento, en minutos, para una muestra aleatoria de 15 programas similares: 12.5, 5.2, 6.8, 3.6, 10.9, 12.8, 7.8, 8.6, 6.3, 6.9, 18.2, 15.4, 9.2, 10.3, 7.3. Supóngase que el tiempo total de procesamiento está modelado, en forma adecuada, por una distribución gama con $\alpha = 3$.

a) Obtener el estimador de máxima verosimilitud para el parámetro de escala θ.
b) El método de momentos, ¿daria un estimado diferente de θ al determinado en el inciso *a*?
c) Mediante la respuesta del inciso *a*), calcular la probabilidad de que el tiempo de procesamiento sea mayor a 20 minutos.

8.21. Un fabricante de fibras sintéticas desea estimar la tensión de ruptura media de una fibra. Diseña un experimento en el que se observan las tensiones de ruptura, en libras, de 16 hilos del proceso seleccionados aleatoriamente. Las tensiones son 20.8, 20.6, 21.0, 20.9, 19.9, 20.2, 19.8, 19.6, 20.9, 21.1, 20.4, 20.6, 19.7, 19.6, 20.3 y 20.7. Supóngase que la tensión de ruptura de una fibra se encuentra modelada por una distribución normal con desviación estándar de 0.45 libras. Construir un intervalo de confianza estimado del 98% para el valor real de la tensión de ruptura promedio de la fibra.

8.22. Con referencia al ejercicio 8.21, ¿cuáles de las siguientes proposiciones son apropiadas papra la interpretación del intervalo de confianza?

a) En la probabilidad de que la tensión promedio verdadera se encuentre, los límites de confianza son de 0.98.
b) Aproximadamente el 98%, de todos los intervalos de confianza calculados con base en repetidas muestras de tamaño, 16 obtenidas en el proceso de fabricación de las fibras incluirán el verdadero valor promedio de la tensión de ruptura.
c) La probabilidad de que la tensión de ruptura para cualquier fibra se encuentre fuera de los límites de confianza es 0.02.

8.23. Mediante el empleo de los métodos de la sección 5.9, genere 100 muestras, cada una de tamaño 16, de una distribución normal con media 100 y desviación estándar 10. Para cada muestra, constrúyase un intervalo de confianza del 95% para μ. ¿Cuántos de estos intervalos contienen el verdadero valor de 100 para μ? Véase el ejercicio 8.36.

8.24. Una tienda de donas se interesa en estimar su volumen de ventas diarias. Supóngase que el valor de la desviación estándar es de $50.

a) Si el volumen de ventas se encuentra aproximado por una distribución normal, ¿cuál debe ser el tamaño de la muestra para que con una probabilidad de 0.95 la media muestral se encuentre a no más de $20 del verdadero volumen de ventas promedio?

b) Si no es posible suponer que la distribución es normal, obtener el tamaño necesario de la muestra para la pregunta *a*.

8.25. Con referencia al ejercicio 8.24, generar 100 muestras, cada una de tamaño igual al determinado en el inciso *a*, de una distribución normal con media y desviación estándar iguales a 400 y 50, respectivamente. Calcular la media muestral para cada muestra. ¿Cuántas medias muestrales se encuentran a no más de $20 del valor conocido de μ? ¿Está su respuesta de acuerdo con lo que se esperaba?

8.26. Se piensa que la diferencia entre el sueldo más bajo y el más alto que se paga por hora a los mecánicos de automóviles es de $9. Si se supone que estos sueldos se encuentran, en forma aproximada, distribuidos según un modelo normal, ¿cuál debe ser el tamaño de la muestra para que con una probabilidad de 0.99 la media muestral se encuentre a no más de un dólar del verdadero salario por hora promedio? Contéstese la misma pregunta sin suponer una distribución normal.

8.27. La Cámara de Comercio de una ciudad se encuentra interesada en estimar la cantidad promedio de dinero que gasta la gente que asiste a convenciones, calculando comidas, alojamiento y entretenimiento por día. De las distintas convenciones que se llevan a cabo en la ciudad, se seleccionaron 60 personas y se les preguntó la cantidad que gastaban por día. Se obtuvo la siguiente información en dólares: 150, 175, 163, 148, 142, 189, 135, 174, 168, 152, 158, 184, 134, 146, 155, 163. Si se supone que la cantidad de dinero gastada en un día es una variable aleatoria distribuida normal, obtener los intervalos de confianza estimados del 90, 95 y 98% para la cantidad promedio real.

8.28. Con referencia al ejercicio 8.21, determinar el intervalo de confianza estimado del 98% para la tensión de ruptura promedio sin suponer que se conoce la desviación estándar de la población. ¿Cómo es este intervalo comparado con el que se obtuvo en el ejercicio 8.21?

8.29. Para verificar la sensititividad de la distribución t de Student con respecto a la suposición de que se muestrea una distribución normal, generar 100 muestras aleatorias cada una de tamaño 10 de una distribución exponencial con $\theta = 20$. Para cada muestra, construir un intervalo de confianza estimado del 95% para la media. ¿Cuántos de estos intervalos contienen el valor medio conocido de $\theta = 20$? Repetir el proceso incrementando el tamaño de la muestra a 30. ¿Existe alguna diferencia? Formular un comentario con respecto a sus resultados. Véase el ejercicio 8.37.

8.30. Una muestra aleatoria de los salarios por hora para nueve mecánicos de automóviles proporcionó los siguientes datos (en dólares): 10.5, 11, 9.5, 12, 10, 11.5, 13, 9, 8.5. Bajo la suposición de que el muestreo se llevó a cabo sobre una población distribuida normal, construir los intervalos de confianza estimados del 90, 95 y 99% para los salarios por hora promedio para todos los mecánicos. Interpretar los resultados.

8.31. Dos universidades financiadas por el gobierno tienen métodos distintos para inscribir a sus alumnos a principios de cada semestre. Las dos desean comparar el tiempo prome-

dio que les toma a los estudiantes completar el trámite de inscripción. En cada universidad se anotaron los tiempos de inscripción para 100 alumnos seleccionados al azar. Las medias y las desviaciones estándares muestrales son las siguientes:

$$\bar{x}_1 = 50.2 \qquad \bar{x}_2 = 52.9$$

$$s_1 = 4.8 \qquad s_2 = 5.4$$

Si se supone que el muestreo se llevó a cabo sobre dos poblaciones distribuidas normales e independientes, obtener los intervalos de confianza estimados del 90, 95 y 99% para la diferencia entre las medias del tiempo de inscripción para las dos universidades. Con base en esta evidencia, ¿se estaría inclinando a concluir que existe una diferencia real entre los tiempos medios para cada universidad?

8.32. Cierto metal se produce, por lo común, mediante un proceso estándar. Se desarrolla un nuevo proceso en el que se añade una aleación a la producción del metal. Los fabricantes se encuentran interesados en estimar la verdadera diferencia entre las tensiones de ruptura de los metales producidos por los dos procesos. Para cada metal se seleccionan 12 especímenes y cada uno de éstos se somete a una tensión hasta que se rompe. La siguiente tabla muestra las tensiones de ruptura de los especímenes en kilogramos por centímetro cuadrado:

Proceso estándar	428	419	458	439	441	456	463	429	438	445	441	463
Proceso nuevo	462	448	435	465	429	472	453	459	427	468	452	447

Si se supone que el muestreo se llevó a cabo sobre dos distribuciones normales e independientes con varianzas iguales, obtener los intervalos de confianza estimados del 90, 95 y 99% para $\mu_S - \mu_N$. Con base en los resultados, ¿se estaría inclinado a concluir que existe una diferencia real entre μ_S y μ_N?

8.33. En dos ciudades se llevó a cabo una encuesta sobre el costo de la vida para obtener el gasto promedio en alimentación en familias constituidas por cuatro personas. De cada ciudad se seleccionó aleatoriamente una muestra de 20 familias y se observaron sus gastos semanales en alimentación. Las medias y las desviaciones estándares muestrales fueron las siguientes:

$$\bar{x}_1 = 135 \qquad \bar{x}_2 = 122$$

$$s_1 = 15 \qquad s_2 = 10$$

Si se supone que se muestrearon dos poblaciones independientes con distribución normal cada una, y varianzas iguales, obtener los intervalos de confianza estimados del 95 y 99% para $\mu_1 - \mu_2$. ¿Se estaría inclinado a concluir que existe una diferencia real entre μ_1 y μ_2?

8.34. Se espera tener una cierta variación aleatoria nominal en el espesor de las láminas de plástico que una máquina produce. Para determinar cuándo la variación en el espesor se encuentra dentro de ciertos límites, cada día se seleccionan en forma aleatoria 12 láminas de plástico y se mide en milímetros su espesor. Los datos que se obtuvieron son los siguientes: 12.6, 11.9, 12.3, 12.8, 11.8, 11.7, 12.4, 12.1, 12.3, 12.0, 12.5, 12.9. Si se supone que el espesor es una variable aleatoria distribuida normal, obtener los intervalos

de confianza estimados del 90, 95 y 99% para la varianza desconocida del espesor. Si no es aceptable una varianza mayor de 0.9 mm, ¿existe alguna razón para preocuparse con base en esta evidencia?

8.35. Mediante el uso de los datos del ejercicio 8.27, obtener un intervalo de confianza estimado del 95% para la varianza desconocida e interpretar el resultado.

8.36. Con referencia al ejercicio 8.23, construir para cada muestra un intervalo de confianza del 95% para σ^2. ¿Cuántos de estos intervalos contienen el valor conocido de 100 para σ^2?¿Este resultado está de acuerdo con lo que se esperaba?

8.37. Para verificar la sensitividad de la distribución chi-cuadrada con respecto a la suposición de que se muestrea una distribución normal, repetir el ejercicio 8.29 construyendo para cada muestra un intervalo de confianza estimado del 95% para σ^2. En relación con los dos tamaños de las muestras, ¿cuántos de estos intervalos contienen el valor conocido de $\sigma^2 = 400$? Con base en estos resultados, comparar las sensitividades de las distribuciones t de Student y chi-cuadrada con respecto a la hipótesis de un muestreo que se lleva a cabo sobre una distribución normal.

8.38. Una agencia estatal tiene la responsabilidad de vigilar la calidad del agua para la cría de peces con fines comerciales. Esta agencia se encuentra interesada en comparar la variación de cierta sustancia tóxica en dos estuarios cuyas aguas se encuentran contaminadas por desperdicios industriales provenientes de una zona industrial cercana. En el primer estuario se seleccionan 11 muestras y en el segundo 8, las cuales se enviaron a un laboratorio para su análisis. Las mediciones en ppm que se observaron en cada muestra se exponen en la tabla 8.6. Si se supone que el muestreo se hizo sobre dos poblaciones independientes distribuidas normales, obtener un intervalo de confianza estimado del 95% para el cociente de las dos varianzas no conocidas σ_1^2/σ_2^2. Con base en este resultado, ¿se podría concluir que las dos varianzas son diferentes? ¿Por qué?

8.39. Con referencia al ejercicio 8.32, construir un intervalo de confianza estimado del 99% para el cociente σ_1^2/σ_2^2, en donde σ_1^2 es la varianza del proceso estándar y σ_2^2 es la varianza del nuevo proceso. Con base en este resultado, ¿es razonable la suposición de que las varianzas son iguales?

8.40. La lista electoral final en una elección reciente para senador, reveló que 1 400 personas

TABLA 8.6 Niveles de una sustancia tóxica (ppm)

Estuario 1	*Estuario 2*
10	11
10	8
12	9
13	7
9	10
8	8
12	8
12	10
10	
14	
8	

de un total de 2 500 seleccionadas aleatoriamente, tienen preferencia por el candidato A con respecto al candidato B.

a) Obtener un intervalo de confianza unilateral inferior del 99% para la verdadera proporción de votantes a favor del candidato A. Con base en este resultado, ¿podría usted afirmar que es probable que A gane la elección? ¿Por qué?

b) Supóngase que se selecciona aleatoriamente una muestra de 225 personas con la misma proporción muestral a favor del candidato A. ¿Son los resultados diferentes a los del inciso *a*)?

c) En este caso, ¿son razonables las suposiciones para los intervalos de confianza aproximados del 99%?

8.41. Se recibe un lote muy grande de artículos proveniente de un fabricante que asegura que el porcentaje de artículos defectuosos en la producción es del 1%. Al seleccionar una muestra aleatoria de 200 artículos y después de inspeccionarlos, se descubren 8 defectuosos. Obtener los intervalos de confianza aproximados del 90, 95 y 99% para la verdadera proporción de artículos defectuosos en el proceso de manufactura del fabricante. Con base en estos resultados, ¿qué se puede concluir con respecto a la afirmación del fabricante?

8.42. Un médico investigador desea estimar la proporción de hombres, en edad madura, que fuman en exceso y que desarrollarán cáncer pulmonar en los siguientes cinco años. El investigador desea seleccionar una cierta cantidad de hombres que hayan fumado por lo menos dos cajetillas de cigarros al día durante 20 años y observarlos durante los próximos cinco años para saber cuántos desarrollan cáncer pulmonar. ¿Cuál debe ser el tamaño de la muestra que el investigador debe seleccionar de manera tal que con una probabilidad de 0.95, la proporción muestral se encuentre a no más de 0.02 unidades de la proporción verdadera?

8.43. Las compañías de auditoría generalmente seleccionan una muestra aleatoria de los clientes de un banco y verifican los balances contables reportados por el banco. Si una compañía de este tipo se encuentra interesada en estimar la proporción de cuentas para las cuales existe una discrepancia entre el cliente y el banco, ¿cuántas cuentas deberán seleccionarse de manera tal que con una confiabilidad del 99% la proporción muestral se encuentre a no más de 0.02 unidades de la proporción real?

8.44. El volumen semanal de ventas de una tienda de descuentos se encuentra representado, en forma adecuada, por una distribución normal con media desconocida μ, pero con una desviación estándar de $\sigma = \$2\ 000$. Debido a muchas influencias de índole menor, se cree que el volumen de ventas semanal promedio puede considerarse como una variable aleatoria. Supóngase que se está pensando asignar una distribución normal a la media semanal con $\mu_0 = \$20\ 000$ y $\sigma_0 = \$200$. Una muestra aleatoria de 16 semanas revela un volumen de ventas promedio muestral de \$21 500.

a) Para una función de pérdida de error cuadrático, obtener el estimador Bayes de μ.

b) Obtener un intervalo estimado de probabilidad bayesiano del 95% para μ.

c) Obtener un intervalo de confianza del 95% para μ y compararlo con el intervalo estimado en el inciso *b)*.

d) Repetir los incisos *a, b* y *c* con $\sigma_0 = 100$. Comentar los resultados.

e) Repetir los incisos *a, b* y *c* con $\sigma_0 = 800$. Comentar los resultados.

f) Supóngase que $n = 64$; asumiendo que $\bar{x} = 21\ 500$, ¿de qué forma afectarían los cambios anteriores las respuestas dadas para los incisos *a, b* y *c*?

8.45. Una oficina estatal determinó que el número de llamadas telefónicas que recibe es una variable aleatoria de Poisson. Debido a las condiciones del mercado, la oficina ha llegado a la conclusión de que el parámetro de Poisson es una variable aleatoria con distribución gama y parámetros de forma y escala iguales a 20 y 4, respectivamente. En un día, seleccionado al azar, se reciben 90 llamadas telefónicas.

a) Para una función de pérdida de error cuadrático, obtener el estimador Bayes del parámetro de Poisson.

b) Obtener un intervalo de probabilidad bayesiano del 95%. (Sugerencia: empléese (5.51).)

8.46. Una compañía constructora de hoteles se encuentra muy interesada en las tensiones de ruptura de los cables de acero que sostendrán un pasillo por encima del vestíbulo del hotel. El contratista hace uso de los servicios de una organización independiente a la cual da las instrucciones necesarias para probar los cables y determinar un límite de tolerancia inferior para la tensión de ruptura de éstos de manera tal que, con una probabilidad de 0.95, el 99% de los cables tenga una tensión de ruptura mayor al límite deseado. La organización selecciona, en forma aleatoria, 20 cables y los prueba para determinar sus tensiones de ruptura. Los resultados de la prueba, en kilogramos por centímetro cuadrado, son 2130, 2158, 2192, 2110, 2145, 2208, 2201, 2195, 2125, 2148, 2166, 2172, 2192, 2138, 2210, 2215, 2108, 2105, 2120 y 2130. Si se supone que la tensión de ruptura es una variable aleatoria distribuida normal, obtener el límite de tolerancia deseado.

8.47. El diámetro interno de un cojinete es una medida crucial en la fabricación de éste. Con base en una muestra aleatoria de 25 cojinetes, la media muestral fue de 3 cm y la desviación estándar muestral fue igual a 0.005 cm. Obtener los límites de tolerancia bilaterales de manera tal que, con una probabilidad de 0.99, el 95% de los diámetros de todos los cojinetes manufacturados por este proceso se encuentren dentro de los límites de tolerancia. Supóngase que el diámetro interno es una variable aleatoria distribuida normal.

8.48. Supóngase que en el ejercicio 8.47 no es posible asumir una distribución normal. Si de los 25 cojinetes, el diámetro más pequeño fue de 2.984 y el más grande de 3.013 y se está interesado en un intervalo que contenga al 90, 95 o 99% de todos los diámetros internos, ¿cuál es la probabilidad que puede asociarse con el intervalo de 2.984 al 3.013 para cada uno de los porcentajes anteriores?

8.49. Supóngase que no es posible asumir una distribución normal en el ejercicio 8.46. Para la misma probabilidad y tamaño muestral, ¿cuál debe ser la proporción de tensiones de ruptura que debe exceder el valor más pequeño de las 20 observaciones? ¿Qué tan grande debe ser la muestra necesaria en este caso para tener la misma probabilidad y proporción del ejercicio 8.46?

8.50. Supóngase que se está muestreando una población cuya distribución de probabilidad es desconocida. ¿Cuál debe ser el tamaño de la muestra necesario para que, con una probabilidad de 0.99, por lo menos el 95% de los valores de la variable aleatoria de interés esté incluido entre los dos valores extremos de la muestra?

8.51. Supóngase que se está muestreando una población cuya distribución de probabilidad es desconocida. ¿Cuál debe ser el tamaño de la muestra necesario para que, con una probabilidad de 0.99, por lo menos el 97% de los valores de la variable aleatoria sea mayor que el valor más pequeño de la muestra?

CAPÍTULO NUEVE

Prueba de hipótesis estadísticas

9.1 Introducción

En el capítulo 8 se examinó la inferencia estadística con respecto a la estimación puntual y por intervalo. En este capítulo se estudiará otra área de la inferencia: la prueba o contraste de una hipótesis estadística. Como se verá, la prueba de una hipótesis estadística tiene una fuerte relación con el concepto de estimación.

Una hipótesis estadística es una afirmación con respecto a alguna característica desconocida de una población de interés. La esencia de probar una hipótesis estadística es el decidir si la afirmación se encuentra apoyada por la evidencia experimental que se obtiene a través de una muestra aleatoria. En forma general, la afirmación involucra ya sea a algún parámetro o a alguna forma funcional no conocida de la distribución de interés a partir de la cual se obtiene una muestra aleatoria. La decisión acerca de si los datos muestrales apoyan estadísticamente la afirmación se toma con base en la probabilidad, y, si ésta es mínima, entonces será rechazada.

En gran medida, el enfoque de este capítulo será más intuitivo que teórico ya que el autor piensa que desde este punto de vista el lector estará en posición de obtener una mejor idea de la esencia de las hipótesis estadísticas. En forma inicial se desarrollarán los fundamentos para la prueba de hipótesis estadísticas. Entonces se examinarán varias áreas de aplicación con respecto a medidas, varianzas y proporciones.

9.2 Conceptos básicos para la prueba de hipótesis estadísticas

Para ilustrar la noción de una hipótesis estadística, supóngase que se tiene interés en el tiempo promedio necesario para terminar una unidad en una línea de armado. Bajo condiciones de operación estándares, el objetivo es tener un tiempo promedio de armado por unidad de 10 minutos. El gerente de la planta decide continuar con el proceso a menos que se encuentre una evidencia sustancial de que el tiempo promedio no es de 10 minutos. La evidencia estará en una muestra aleatoria de tamaño n obtenida de la distribución de interés para el tiempo de armado de una unidad. ¿Cómo debe decidirse si el proceso continúa en operación?

La respuesta a este tipo de preguntas es el principal objetivo del presente capítulo. Nótese que no es de interés, *per se*, la estimación del tiempo medio desconocido μ, sino determinar si el valor de μ es 10. En otras palabras, antes de que la muestra se obtenga, ya se ha conjeturado que el muestreo se llevará a cabo sobre una distribución cuya media es 10. Si la afirmación es estadísticamente plausible con base en la evidencia experimental, entonces se asumirá que el valor promedio objetivo es de 10 minutos y, por lo tanto, se dejará que el proceso continúe. Por otro lado, si la afirmación no está apoyada estadísticamente por la evidencia muestral, el gerente de la planta puede decidir detener el proceso para llevar a cabo los ajustes necesarios.

A la afirmación de que $\mu = 10$ se le llama *hipótesis nula* y se escribirá como:

$$H_0: \mu = 10.$$

Nótese que con H_0 se ha especificado un solo valor para el parámetro en cuestión. De hecho, si una hipótesis estadística asigna valores particulares a todos los parámetros desconocidos e identifica la forma funcional de la distribución de interés, recibe el nombre de *hipótesis sencilla o simple*. De otra forma, se conoce como *hipótesis compuesta*. De esta manera, H_0: $\mu = 10$ es una hipótesis sencilla sólo si se especificaron la forma funcional de la distribución de interés y los valores de los parámetros desconocidos (si es que los hay). Si la hipótesis nula se hubiese propuesto como H_0: $\mu \leq 10$ o H_0: $\mu \geq 10$, ésta no sería una hipótesis simple ya que no asigna ningún valor específico para μ.

Una hipótesis nula debe considerse como verdadera a menos que exista suficiente evidencia en contra. En otras palabras, se rechazará la hipótesis nula de que el tiempo promedio de armado es de 10 minutos, sólo si la evidencia experimental se encuentra muy en contra de esta afirmación. Un paralelo muy cercano a esta interpretación es el de los procesos judiciales en los que el acusado es inocente hasta que no se demuestre lo contrario. Esto es, definiendo a la hipótesis nula como "inocente", se insiste en que se rechazará sólo si el juicio proporciona evidencia suficiente en contra de ésta.

A continuación se analizan las posibles decisiones que pueden tomarse con respecto a la hipótesis nula H_0: $\mu = 10$. Al hacer esto deben tomarse en cuenta las consecuencias que pueden originarse como resultado del verdadero estado de la naturaleza: en realidad μ, puede o no ser igual a 10. En forma sencilla, existen dos posibles decisiones con respecto a H_0 *(rechazar H_0 o equivocarse al rechazar H_0)*.* Sin embargo, cada una de estas decisiones tiene las siguientes dos consecuencias con respecto al estado de la naturaleza:

$$\text{Rechazar } H_0 \begin{cases} \text{cuando de hecho } H_0 \text{ es cierta} \\ \\ \text{cuando de hecho } H_0 \text{ es falsa} \end{cases} \qquad \begin{matrix}\text{Equivocarse} \\ \text{al rechazar } H_0\end{matrix} \begin{cases} \text{cuando de hecho } H_0 \text{ es cierta} \\ \\ \text{cuando de hecho } H_0 \text{ es falsa} \end{cases}$$

Si la decisión es el rechazar a H_0, entonces puede que se rechace algo que es cierto (decisión incorrecta) o que se rechace algo que en realidad es falso (decisión

* La razón de por qué se ha usado la frase "equivocarse al rechazar H_0" más que "aceptar H_0" será evidente más adelante.

correcta). Si no se puede rechazar H_0, entonces no puede rechazarse algo que es cierto (decisión correcta), o no puede rechazarse algo que en realidad es falso (decisión incorrecta). Por lo tanto, si la decisión es rechazar o equivocarse al rechazar H_0, existen dos posibilidades de tomar una decisión equivocada con respecto al verdadero estado de la naturaleza.

Cuando se toma una decisión con respecto a una hipótesis nula, dos de las posibles consecuencias relativas al verdadero estado de la naturaleza conducen a errores inferenciales. El rechazo de la hipótesis H_0 cuando en realidad H_0 es cierta, constituye lo que se denomina *error de tipo I*. Equivocarse al rechazar H_0 cuando en realidad H_0 es falsa, constituye lo que se denomina *error de tipo II*. El lector debe notar que sólo es posible el error de tipo I cuando la decisión es el rechazar la hipótesis nula, mientras que el error de tipo II sólo es posible cuando la decisión es el no rechazar H_0. En otras palabras, si la hipótesis nula realmente es cierta, sólo puede cometerse un error de tipo I; si la hipótesis nula es falsa, sólo puede cometerse un error de tipo II. No pueden cometerse ambos errores en forma simultánea. De manera obvia, el interés recae en la posiblidad de cometer un tipo, cualquiera, de error. Sin embargo, es importante comprender que una decisión con respecto a una hipótesis estadística es un proceso inferencial, el cual siempre se encuentra sujeto a error. La decisión de rechazar H_0 no necesariamente significa que H_0 sea falsa; pero la evidencia muestral con base en la cual se toma la decisión proporciona un grado de confiabilidad (paralelo al de la estimación de intervalo) con el que puede procederse como si H_0 fuese falsa.

Es necesario tener alguna cantidad que mida la posibilidad de cometer alguno de estos errores. Esta medida es una probabilidad.

Definición 9.1 La probabilidad de rechazar H_0, dado que H_0 es cierta, se define como la probabilidad (o tamaño) del error de tipo I y se denota por α, $0 \leq \alpha \leq 1$.

Definición 9.2 La probabilidad de no rechazar H_0, dado que H_0 es falsa, se define como la probabilidad (o tamaño) del error de tipo II y se denota por β, $0 \leq \beta \leq 1$.

Por lo tanto, las probabilidades de los errores de tipo I y tipo II están dadas por las proposiciones

$$P(\text{rechazar } H_0 \mid H_0 \text{ es cierta}) = \alpha \tag{9.1}$$

y

$$P(\text{no poder rechazar } H_0 \mid H_0 \text{ es falsa}) = \beta. \tag{9.2}$$

Nótese que tanto α como β son probabilidades condicionales. No pueden obtenerse las probabilidades de los errores de tipo I y tipo II en un sentido absoluto, debido a que el estado de la naturaleza no es conocido. Más bien, puede calcularse la probabilidad α de rechazar H_0 sólo si se asume que H_0 es cierta, o la probabilidad β de equivocarse el rechazar H_0, si se asume que H_0 es falsa.

Cuando una afirmación se incorpora en la proposición de la hipótesis nula, se necesita una regla que indique qué decisión tomar con respecto a H_0 una vez que se en-

Test

cuentra disponible la evidencia muestral. Esta regla recibe el nombre de prueba de una hipótesis estadística.

Definición 9.3 Una *prueba* de una hipótesis estadística con respecto a alguna característica desconocida de la población de interés es cualquier regla para decidir si se rechaza la hipótesis nula con base en una muestra aleatoria de la población.

La decisión se basa en alguna estadística apropiada la cual recibe el nombre de *estadística de prueba*. Para ciertos valores de la estadística de prueba, la decisión será el rechazar la hipótesis nula. Estos valores constituyen lo que se conoce como la *región crítica* de la prueba. Por ejemplo, recuérdese la hipótesis nula H_0: $\mu = 10$. Para un tamaño n dado de la muestra, supóngase que se decide rechazar H_0 si se observa un valor de la media muestral $\overline{X}$ que sea más grande que 12. Entonces, $\overline{X}$ es la estadística de prueba, el valor $\overline{X} = 12$ es el *valor crítico*, y el conjunto de valores mayores que 12 constituyen la región crítica de la prueba.

Para mostrar en forma gráfica la región crítica, supóngase que n es suficientemente grande de manera tal que la distribución de muestreo de la estadística de prueba $\overline{X}$, dado que H_0 es cierta, es esencialmente una distribución normal. La figura 9.1 muestra la región crítica como el área sombreada a la derecha del valor crítico $\overline{X} = 12$. El área de la región crítica es igual al tamaño del error de tipo I. En otras palabras, $P(\overline{X} > 12|\mu = 10) = \alpha$. La interpretación de α es análoga a la de los intervalos de confianza. Esto es, la probabilidad α es sólo una referencia con respecto a la región $\overline{X} > 12$ involucrando a la variable aleatoria $\overline{X}$, dado que $\mu = 10$. Pero la decisión de rechazar H_0 se tomará con base en una sola muestra de tamaño n, a partir de la cual se calculará el estimador de $\bar{x}$. De esta forma, si $\bar{x} > 12$, esto no significa que la probabilidad de que H_0 sea correcta es α; más bien, esto implica una interpretación de frecuencia para α cuando se toman muchas muestras. En otras palabras, si el valor de μ es realmente 10, y si se tomasen en forma repetida muestras de tamaño n de la población, debe esperarse que en un $100\alpha\%$ de las veces, se encuentre un valor de la estadística de prueba $\overline{X}$ mayor que 12, y de esta forma debe

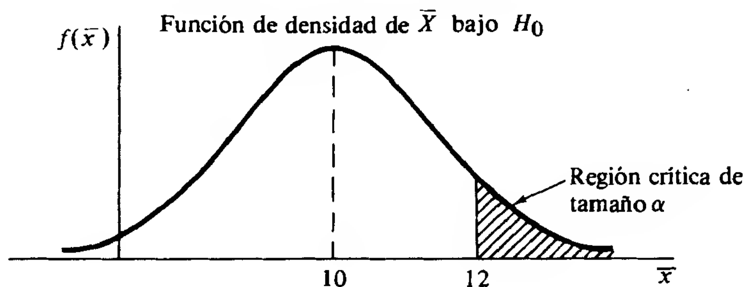


FIGURA 9.1 La región crítica como un área

rechazarse la hipótesis nula. Sólo en este sentido puede decirse que la confiabilidad al rechazar H_0, cuando el estimador $\overline{X} > 12$ es igual al complemento del error α de tipo I, o $1 - \alpha$.

Para construir una regla de decisión apropiada en la prueba de una hipótesis estadística, también es necesario establecer una *hipótesis alternativa* que refleje el valor posible o intervalo de valores del parámetro de interés si la hipótesis nula es falsa. Esto es, la hipótesis alternativa representa alguna forma de negación de la hipótesis nula. Generalmente la hipótesis alternativa se representa por H_1 y puede ser simple o compuesta. A pesar de que no se pretende una generalización, en muchas ocasiones es deseable establecer una hipótesis nula que sea más específica que la alternativa. De esta manera, la hipótesis nula es simple en forma general, mientras que la alternativa es una hipótesis compuesta. Por ejemplo, supóngase que el gerente de la planta sospecha que el tiempo de armado promedio es mayor de 10 minutos. Entonces las hipótesis nula y alternativa apropiadas podrían ser

$$H_0: \mu = 10,$$

$$H_1: \mu > 10.$$

La razón de ello es que si la evidencia muestral no apoya el rechazo de la hipótesis nula, entonces el gerente de la planta podría proceder como si H_0 fuese cierta. De otra manera, la sospecha podría justificarse y entonces puede ser necesario emprender alguna acción para corregir la falla.

De acuerdo con la definición 9.3, el procedimiento de prueba se construye de manera tal que la hipótesis nula sea o no rechazada. En este sentido, se dice que H_0 es la hipótesis a ser probada. Sin embargo, con la inclusión de la hipótesis alternativa, puede ser más descriptivo decir que probar una hipótesis estadística es proporcionar una decisión entre H_0 y H_1. Por ello debe ejercerse una precaución extrema al establecer las hipótesis nula y alternativa.

Se regresará a la analogía del proceso judicial para proporcionar una idea más clara sobre la materia. Si la hipótesis nula es "inocente", entonces, con toda seguridad, la hipótesis alternativa es "culpable". El rechazo de la hipótesis nula implicaría que el juicio ha sido capaz de proporcionar suficiente evidencia para garantizar un veredicto de culpable. Por otro lado, si el juicio no presenta evidencia sustancial, el veredicto será inocente. Esta decisión no implica necesariamente que el acusado sea inocente, más bien hace énfasis en la falta de evidencia sustancial necesaria para condenar al acusado. Por lo tanto, en cierto sentido un veredicto de culpable (el rechazo de H_0) debe considerarse como una decisión más fuerte que un veredicto de inocente (equivocación al rechazar H_0), lo cual surge del principio judicial generalmente aceptado de que es peor condenar a una persona inocente que dejar ir a una culpable. Si el veredicto es culpable, se desea tener un grado muy alto de seguridad de que no se va a condenar a una persona inocente. Por lo tanto, en muchas situaciones el error de tipo I se considera como un error mucho más grave que el error de tipo II.

En la prueba de hipótesis estadísticas el enfoque general es aceptar la premisa de que el error de tipo I es mucho más serio que el error de tipo II, y formular las hi-

pótesis nula y alternativa de acuerdo con lo anterior. Como resultado se tiene que muchas veces se selecciona con anticipación el tamaño máximo del error de tipo I que puede tolerarse y se intenta construir un procedimiento de prueba que minimice el tamaño del error de tipo II. En otras palabras, no es posible fijar tanto a α como a β y diseñar alguna regla de decisión para probar H_0 contra H_1, dada una muestra aleatoria de tamaño n. Es por esta razón que se dice "equivocación al rechazar H_0" más que "aceptar H_0" cuando la evidencia muestral no apoya el rechazo de la hipótesis nula.

Un principio sencillo y razonable al obtener reglas de decisión para la prueba de hipótesis estadísticas es seleccionar aquel procedimiento de prueba que tenga el tamaño más pequeño para el error de tipo II entre todos los procedimientos que tengan el mismo tamaño para el error de tipo I. En este contexto debe notarse que el valor de α no puede hacerse muy pequeño sin que se incremente el valor de β. En otras palabras, para una muestra de tamaño n dado, el tamaño del error de tipo II normalmente aumentará conforme el tamaño del error de tipo I disminuya. Lo que, en forma general, se hace en la práctica, es ajustar el tamaño del error de tipo I cambiando el valor crítico de la estadística de prueba para así alcanzar un balance satisfactorio entre los tamaños de los dos errores. Sin embargo, cuando se hace esto debe tenerse en mente el máximo tamaño del error de tipo I que puede tolerarse en una situación en particular. Por ejemplo, recuérdese de nuevo la hipótesis nula H_0: $\mu = 10$ contra la hipótesis alternativa H_1: $\mu > 10$. Entonces β es igual a la probabilidad de equivocarse al rechazar H_0 cuando H_1 es cierta. Al igual que antes, sea $\bar{X}$ la estadística de prueba. La figura 9.2 muestra cómo, mediante el cambio del valor crítico de 12 a 11, el tamaño de error de tipo I disminuye (éste se encuentra por debajo de la curva que está a la izquierda en ambos casos), pero crece el tamaño del error de tipo II (éste se muestra bajo la curva que se encuentra a la derecha en ambos casos).

La probabilidad α del error de tipo I también se conoce como el *nivel de significancia estadístico*. En este contexto la palabra "significancia" sólo implica que la

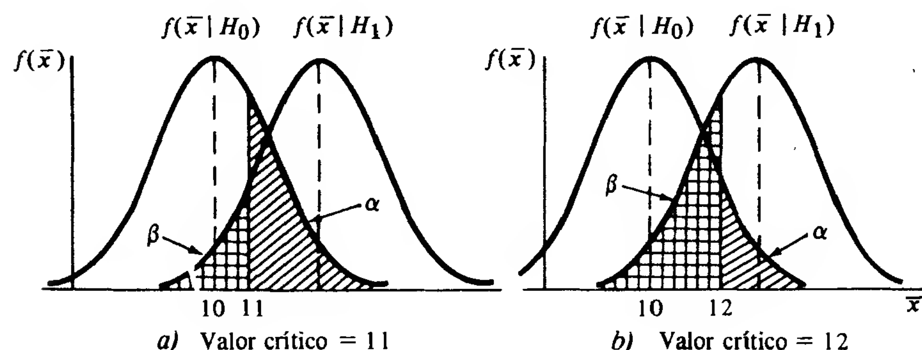


FIGURA 9.2 El efecto sobre α y β al cambiar el valor crítico

evidencia muestral es tal que garantiza el rechazo de H_0 a un nivel dado de α. En consecuencia, la frase "El rechazo de H_0 es *estadísticamente discernible* a un nivel dado α", es más apropiada. Un ejemplo ilustrará los conceptos anteriores.

Ejemplo 9.1 Supóngase que puede tolerarse un tamaño del error de tipo I hasta de 0.06 cuando se prueba la hipótesis nula

$$H_0: \mu = 10$$

contra la hipótesis alternativa

$$H_1: \mu > 10$$

para el problema del tiempo de armado. Supóngase que la distribución del tiempo necesario para armar una unidad es normal con desviación estándar $\sigma = 1.4$ minutos. Se observan los tiempos de armado de 25 unidades seleccionadas aleatoriamente y escoge la media muestral $\overline{X}$ como la estadística de prueba. En particular, se desea comparar las siguientes regiones críticas.

Prueba A: Rechazar H_0 si $\overline{X} > 10.65$

Prueba B: Rechazar H_0 si $\overline{X} > 10.45$

Prueba C: Rechazar H_0 si $\overline{X} > 10.25$

para determinar cuál de éstas satisface el tamaño del error de tipo I que puede tolerarse y cuál tiene el valor más pequeño de β entre las tres.

Para determinar la probabilidad del error de tipo I, se asumirá que H_0 es cierta y se calculará

$$P(\overline{X} > c \mid \mu = 10) = \alpha,$$

en donde c es el valor crítico, o frontera de la región crítica. Ya que se asume que el muestreo se lleva a cabo sobre una distribución normal, bajo H_0, $\overline{X} \sim N(10, 1.4/\sqrt{25})$. Por lo tanto, para la prueba A

$$\begin{aligned}\alpha &= P(\overline{X} > 10.65 \mid \mu = 10)\\ &= P[Z > (10.65 - 10)/0.28 \mid \mu = 10]\\ &= P(Z > 2.32 \mid \mu = 10)\\ &= 0.0102.\end{aligned}$$

De manera similar, para la prueba B

$$\alpha = P(\overline{X} > 10.45 \mid \mu = 10) = P(Z > 1.61 \mid \mu = 10) = 0.0537,$$

y para la prueba C

$$\alpha = P(\overline{X} > 10.25 \mid \mu = 10) = P(Z > 0.89 \mid \mu = 10) = 0.1867.$$

Nótese que el tamaño del error de tipo I para la prueba C es mayor al límite impuesto de 0.06, mientras que para las pruebas A y B, éste es menor que el límite dado. Puesto que C no reúne los requisitos, no será ya considerada.

Ya que ni la prueba A ni la B han violado el tamaño máximo del error de tipo I, se determinará cuál de estas dos tiene el tamaño más pequeño para el error de tipo II. Recuérdese que la ocurrencia de un error de tipo II implica que H_0 es falsa. Entonces, para un tamaño de la muestra y un valor máximo de α dados, el tamaño del error del tipo II será, en forma estricta, una función del intervalo de valores del parámetro desconocido como se encuentran especificados en la hipótesis alternativa. En otra palabras

$$\beta(\mu) = P(\overline{X} \leq c \mid \mu > 10).$$

En particular, supóngase que el valor real de μ es igual a 10.4. Entonces, para la prueba A

$$\beta(10.4) = P(\overline{X} \leq 10.65 \mid \mu = 10.4) = P(Z \leq 0.89 \mid \mu = 10.4) = 0.8133,$$

mientras que para la prueba B

$$\beta(10.4) = P(\overline{X} \leq 10.45 \mid \mu = 10.4) = P(Z \leq 0.18 \mid \mu = 10.4) = 0.5714.$$

De esta forma, si $\mu = 10.4$, la probabilidad de que la prueba A se equivoque al rechazar la hipótesis nula de que $\mu = 10$ es de 0.8133, y la correspondiente probabilidad para la prueba B es de 0.5714. Para este valor particular de la hipótesis alternativa, la prueba B es mejor que la A.

Al ilustrar el intervalo de valores de las probabilidades β para estas dos pruebas, se continúa el proceso de calcular el tamaño del error de tipo II para otros valores representativos. En la tabla 9.1 se da la información pertinente. Posteriormente se ilustrará que para una hipótesis alternativa dada y un tamaño fijo del error de tipo I, puede reducirse el tamaño del error de tipo II mediante el incremento del tamaño de la muestra.

Con base en la información proporcionada en la tabla 9.1, pueden formularse las siguientes observaciones. Conforme el tamaño del error de tipo I disminuye (prueba A), el tamaño del error de tipo II aumenta. Si la afirmación propuesta por la hipótesis nula es falsa pero difiere muy poco del verdadero valor, la opción de no rechazar H_0 es alta. Sin embargo, si la hipótesis nula es falsa por una cantidad muy grande, la probabilidad de equivocarse al detectar su falsedad es pequeña. De esta forma, al comparar las pruebas A y B, si puede tolerarse un tamaño del error de tipo I hasta de 0.06, entonces la prueba B es mejor que la A debido a que sus probabilidades β son, de manera uniforme, más pequeñas que las de la prueba A.

TABLA 9.1 Probabilidades para el error de tipo II para las pruebas A y B

μ	10.2	10.4	10.6	10.8	11.0	11.2	11.4
Prueba A	0.9463	0.8133	0.5714	0.2946	0.1056	0.0250	0.0037
Prueba B	0.8133	0.5714	0.2946	0.1056	0.0250	0.0037	0.0003

9.3 Tipos de regiones críticas y la función de potencia

Con anterioridad se sugirió que es deseable establecer una hipótesis nula simple. De hecho, también es deseable establecer una hipótesis alternativa simple ya que sólo en este caso es posible determinar valores únicos de los tamaños de los errores tipo I y tipo II. Con el propósito de ilustrar lo anterior, recuérdese el ejemplo 9.1. Supóngase que para éste también se ha formulado la siguiente hipótesis alternativa H_1: $\mu = 10.8$. Entonces para las pruebas A y B, los tamaños de los errores de tipo I permanecerán en 0.0102 y 0.0537, respectivamente. Pero en este caso la probabilidad del error de tipo II para cualquiera de las pruebas tendrá un solo valor más que un intervalo de valores, como en el ejemplo 9.1. Sin embargo, debe notarse que una hipótesis alternativa simple puede tener una aplicación real limitada. De acuerdo con lo anterior, se procederá bajo la hipótesis de que la hipótesis nula es simple y la alternativa compuesta.

En este contexto se desean estudiar los tipos de regiones críticas que pueden surgir. Considérese la hipótesis nula simple.

$$H_0: \theta = \theta_0$$

con respecto al parámetro de interés θ, cuando se muestrea una distribución cuya función de densidad de probabilidad es $f(x; \theta)$, en donde θ_0 es el valor propuesto de θ. Si la hipótesis alternativa es de la forma.

$$H_1: \theta > \theta_0$$

o

$$H_1: \theta < \theta_0,$$

Se dice que H_1 es una *hipótesis alternativa unilateral*, debido a que los posibles valores de θ bajo H_1 se encuentran a un lado del valor propuesto bajo H_0. La región crítica también recibe el nombre de región de rechazo unilateral debido a que es, en forma intuitiva, razonable rechazar H_0 para los valores de una estadística de prueba apropiada que, si H_0 fuese cierta, son extremos en la dirección que especifica la hipótesis alternativa. Vale la pena notar que la hipótesis alternativa debe formularse sólo si el valor de uno de los parámetros que se encuentre en el lado opuesto, no tiene sentido para el investigador. De otro modo, debe establecerse una *hipótesis alternativa bilateral*. Esto es, si la hipótesis alternativa no proporciona una dirección con respecto al valor propuesto de θ_0, entonces se dice que H_1 es una hipótesis alternativa bilateral de la forma

$$H_1: \theta \neq \theta_0.$$

Una hipótesis alternativa bilateral implica la existencia de una región crítica bilateral* ya que H_1 incluye valores de θ que se encuentran a ambos lados del valor propuesto de θ_0. Para este caso, la decisión se inclina a rechazar la hipótesis nula para aquellos valores de la estadística de prueba que, si H_0 fuese cierta, son extremos en cualquier dirección.

* En forma general, una región crítica bilateral es simétrica; las dos partes de la región se seleccionan de tal manera que el área bajo cada una de las regiones sea igual.

Potencia de la Prueba

TABLA 9.2 Potencias de las pruebas A y B para el ejemplo 9.1

μ	10.2	10.4	10.6	10.8	11.0	11.2	11.4
Prueba A	0.0537	0.1867	0.4286	0.7054	0.8944	0.9750	0.9963
Prueba B	0.1867	0.4286	0.7054	0.8944	0.9750	0.9963	0.9997

Si se asume una hipótesis alternativa compuesta, es necesario generalizar los medios por los cuales se puede evaluar la interpretación de una prueba dada, en forma especial cuando se compara ésta con otras pruebas. Como se ilustra en el ejemplo 9.1, el tamaño del error de tipo II varía para los diferentes valores de θ de la hipótesis alternativa cuando H_1 es compuesta. De esta forma el tamaño del error de tipo II se obtiene como una función de los valores alternativos de θ bajo H_1. Debe notarse que $\beta(\theta)$ se conoce como la ***función característica de operación,*** y cuando se grafica $\beta(\theta)$ para diversos valores de θ de H_1, se obtiene una curva característica de operación (CO).

Dado que $\beta(\theta)$ es la probabilidad de que un valor de la estadística de prueba no se encuentre en la región crítica cuando H_0 es falsa, entonces $1 - \beta(\theta)$ representa la probabilidad de que un valor de la estadística de prueba se encuentre dentro de la región crítica cuando H_0 es falsa. Esta probabilidad se conoce como la ***función potencia*** de la prueba. En otras palabras, las funciones potencia y características de operación son complementarias.

Definión 9.4 La función $P(\theta) = 1 - \beta(\theta)$ recibe el nombre de función potencia y representa la probabilidad de rechazar la hipótesis nula cuando ésta es falsa; es decir, cuando el valor del parámetro de H_1 es cierto.*

En esencia, la potencia de una prueba es la probabilidad de detectar que H_0 es, en forma verdadera, falsa; de aquí el uso de la palabra "potencia". Como ilustración, recuérdese el ejemplo 9.1. Los complementos de las probabilidades de los errores de tipo II que se encuentran en la tabla 9.1 son las potencias de las pruebas A y B para los valores indicados de μ cuando se prueba H_0: $\mu = 10$ contra H_1: $\mu > 10$. Estos valores se encuentran en la tabla 9.2. De esta información, es evidente que la prueba B es más poderosa que la prueba A. Pueden graficarse las funciones características y de potencia de las pruebas A y B contra los valores de μ, dando las curvas características de operación y de potencia que se ilustran en la figura 9.3.

Recuérdese que para un α fijo y una hipótesis alternativa dada, puede disminuirse el tamaño del error de tipo II si se incrementa el tamaño de la muestra. Por lo tanto, se desprende que la función de potencia aumentará conforme aumenta el tamaño de la muestra. Como ilustración, considérense las pruebas A y B del ejemplo 9.1 para las que el tamaño de la muestra se aumenta hasta un valor de 50. Dado que se insiste que los tamaños del error de tipo I siguen siendo los mismos para las

* Si H_0 es cierta, algunos autores definen la potencia para ser igual al tamaño del error de tipo I.

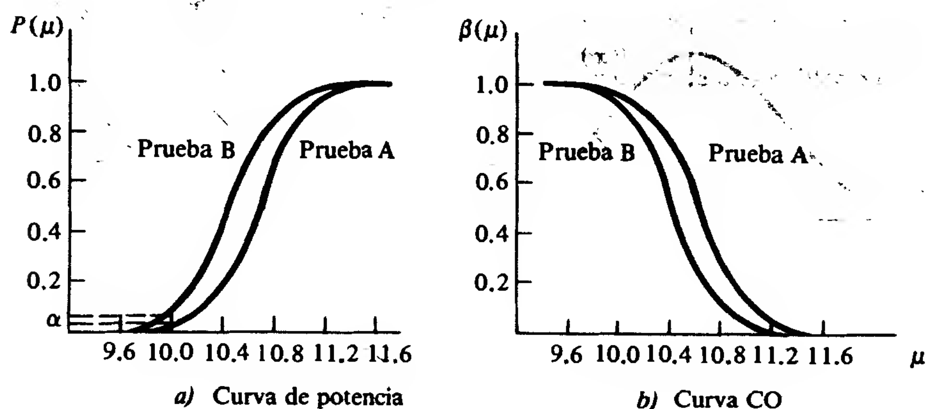


FIGURA 9.3 Comparación de las funciones potencia y característica de operación para A y B

pruebas A y B, sus valores críticos pueden disminuir de valor debido al incremento en el tamaño de la muestra. En particular, pra la prueba A

$$P(\overline{X} > c_A \mid \mu = 10) = 0.0102,$$

o

$$\frac{c_A - 10}{1.4/\sqrt{50}} = 2.32,$$

$$c_A = 10.46.$$

De manera similar, para la prueba B

$$P(\overline{X} > c_B \mid \mu = 10) = 0.0537,$$

y $c_B = 10.32$. La tabla 9.3 contiene información comparable con la de las tablas 9.1 y 9.2 para $n = 50$.

También puede mostrarse la potencia para diferentes valores de μ relativos a la distribución de muestreo de la estadística $\overline{X}$. Considérese, por ejemplo, la prueba B,

TABLA 9.3 Potencias y probabilidades β de las pruebas A y B para $n = 50$

μ		10.2	10.4	10.6	10.8	11.0	11.2	11.4
Prueba A	$P(\mu)$	0.0951	0.3821	0.7611	0.9573	0.9968	0.9999	$\simeq 1$
	$\beta(\mu)$	0.9049	0.6179	0.2389	0.0427	0.0032	0.0001	$\simeq 0$
Prueba B	$P(\mu)$	0.2709	0.6554	0.9207	0.9922	0.9997	$\simeq 1$	$\simeq 1$
	$\beta(\mu)$	0.7291	0.3446	0.0793	0.0078	0.0003	$\simeq 0$	$\simeq 0$

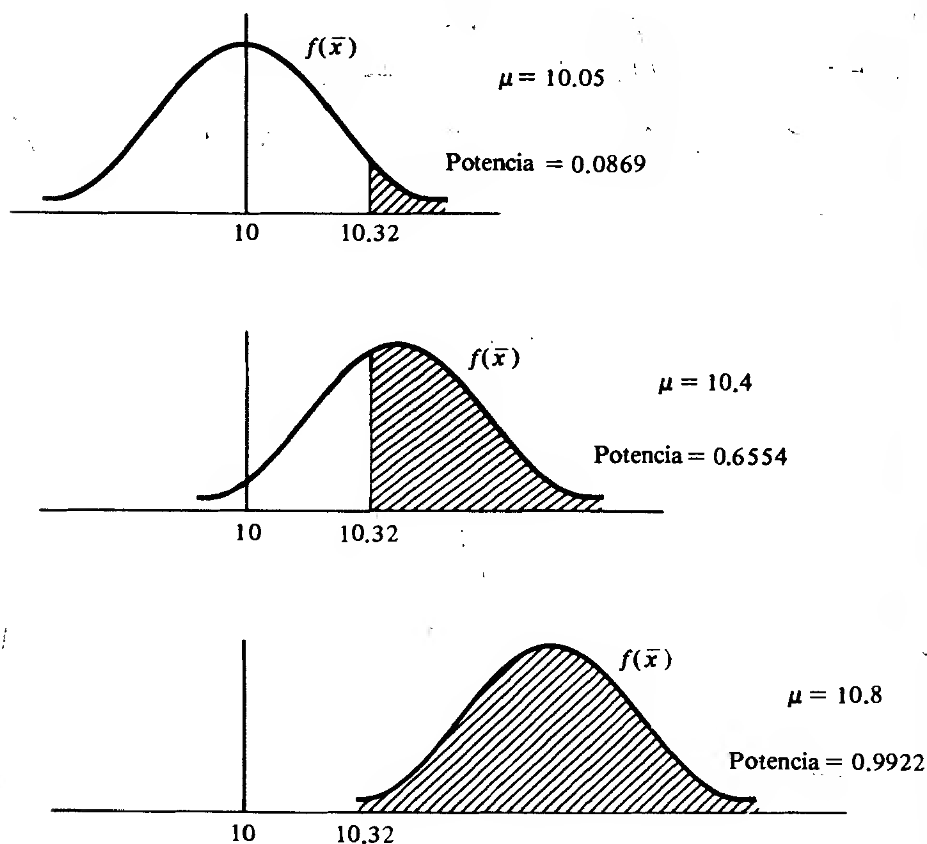


FIGURA 9.4 Probabilidades de rechazo de H_0 para la prueba B ($n = 50$)

en la que el valor crítico es $c_B = 10.32$ para $n = 50$. La figura 9.4 muesra la distribución de $\overline{X}$ para distintos valores de $\mu > 10$, en donde el área sombreada es la potencia o la probabilidad de rechazar H_0. Nótese que conforme el valor de μ se aleja del valor propuesto bajo H_0, la potencia de la prueba aumenta.

9.4 Las mejores pruebas

En la última sección se determinó que la evaluación de la prueba de una hipótesis estadística debe hacerse con base en su función de potencia. En esta sección se regresará al problema igualmente importante de cómo construir una buena prueba. En un sentido teórico, el método para construir buenas pruebas es más claro cuando tanto las hipótesis nula y alternativa son simples o cuando ambas son cumpuestas. En este

punto, se considerará un teorema para construir las mejores pruebas en el caso sencillo de H_0 contra H_1. Este teorema también tiene alguna aplicación en casos más prácticos.

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño n de una población cuya función (densidad) de probabilidad es $f(x; \theta)$, y considérese la hipótesis

$$H_0: \theta = \theta_0$$

contra

$$H_1: \theta = \theta_1,$$

en donde se especifican θ_0 y θ_1. Supóngase que α es el tamaño máximo del error de tipo I que se puede tolerar. Entonces la mejor prueba para H_0 contra H_1 es aquélla que tiene el tamaño más pequeño del error de tipo II (y de esta forma la mayor potencia) de entre todas las pruebas que tengan un tamaño del error de tipo I no mayor que α. Se pueden determinar las regiones críticas para estas pruebas mediante el uso del siguiente teorema, el cual se conoce como lema de Neyman-Pearson:

Teorema 9.1 Si existe una región crítica C de tamaño α y una constante positiva k tal que

$$\frac{L_0(x_1, x_2, \ldots, x_n; \theta_0)}{L_1(x_1, x_2, \ldots, x_n; \theta_1)} \leq k \quad \text{interior } C,$$

$$\frac{L_0(x_1, x_2, \ldots, x_n; \theta_0)}{L_1(x_1, x_2, \ldots, x_n; \theta_1)} \geq k \quad \text{exterior } C,$$

entonces C es la mejor región crítica de tamaño α para probar H_0: $\theta = \theta_0$ contra H_1: $\theta = \theta_1$, en donde L_0 y L_1 son las funciones de verosimilitud relativa a H_0 y H_1, respectivamente.

La demostración del teorema 9.1 se encuentra más allá del alcance de este libro. Sin embargo, puede aclararse la utilidad de este teorema mediante los siguientes ejemplos.

Ejemplo 9.2 Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño n de una distribución normal con media μ desconocida y varianza σ^2 conocida. Determinar la mejor región crítica de tamaño α para probar

$$H_0: \mu = \mu_0$$

contra

$$H_1: \mu = \mu_1,$$

en donde $\mu_1 > \mu_0$.

Bajo H_0 la función de verosimilitud es

$$L_0(x_1, x_2, \ldots, x_n; \mu_0) = (\sqrt{2\pi}\,\sigma)^{-n} \exp\left[-\sum_{i=1}^{n} (x_i - \mu_0)^2 \Big/ 2\sigma^2\right],$$

y bajo H_1 ésta es

$$L_1(x_1, x_2, \ldots, x_n; \mu_1) = (\sqrt{2\pi}\,\sigma)^{-n} \exp\left[-\sum_{i=1}^{n} (x_i - \mu_1)^2 \Big/ 2\sigma^2\right]$$

Entonces, de acuerdo con el teorema 9.1, la mejor región crítica es aquélla para la cual

$$\frac{\exp\left[-\sum(x_i - \mu_0)^2/2\sigma^2\right]}{\exp\left[-\sum(x_i - \mu_1)^2/2\sigma^2\right]} \leqslant k.$$

Esta desigualdad puede escribirse como

$$\exp\left\{\frac{1}{2\sigma^2}\left[\sum(x_i - \mu_1)^2 - \sum(x_i - \mu_0)^2\right]\right\} \leqslant k, \tag{9.3}$$

la cual, después de tomar los logaritmos, se reduce a

$$\sum(x_i - \mu_1)^2 - \sum(x_i - \mu_0)^2 \leqslant 2\sigma^2 \ln(k). \tag{9.4}$$

El lado izquierdo de (9.4) se simplifica de la siguiente manera:

$$\begin{aligned}\sum(x_i - \mu_1)^2 - \sum(x_i - \mu_0)^2 &= \sum x_i^2 - 2\mu_1\sum x_i + n\mu_1^2 - \sum x_i^2 + 2\mu_0\sum x_i - n\mu_0^2 \\ &= n(\mu_1^2 - \mu_0^2) - 2(\mu_1 - \mu_0)\sum x_i.\end{aligned}$$

Sustituyendo en (9.4) se tiene

$$n(\mu_1^2 - \mu_0^2) - 2(\mu_1 - \mu_0)\sum x_i \leqslant 2\sigma^2 \ln(k),$$

o

$$-2(\mu_1 - \mu_0)\sum x_i \leqslant 2\sigma^2 \ln(k) - n(\mu_1^2 - \mu_0^2).$$

Puesto que $\mu_1 > \mu_0$, la cantidad $-2(\mu_1 - \mu_0)$ es negativa; así que

$$\sum x_i \geqslant \frac{n(\mu_1^2 - \mu_0^2) - 2\sigma^2 \ln(k)}{2(\mu_1 - \mu_0)},$$

o

$$\bar{x} \geqslant \frac{n(\mu_1^2 - \mu_0^2) - 2\sigma^2 \ln(k)}{2n(\mu_1 - \mu_0)}. \tag{9.5}$$

La expresión (9.5) define la forma de la mejor región crítica para probar H_0: $\mu = \mu_0$ contra H_1: $\mu = \mu_1$ en donde $\mu_1 > \mu_0$. De manera sencilla, la mejor región crítica es el extremo derecho de la distribución de muestreo de $\bar{X}$ bajo la hipóte-

sis nula. Para un α dado, el valor crítico $\bar{x}_0$ puede encontrarse mediante una elección apropiada de la constante positiva K, de manera tal que

$$P(\bar{X} \geqslant \bar{x}_0 \mid \mu = \mu_0) = \alpha.$$

En particular, supóngase que se escoge un tamaño del error de tipo I igual a 0.05. Entonces el valor crítico de $\bar{x}_0$ es tal que

$$P(\bar{X} \geqslant \bar{x}_0 \mid \mu = \mu_0) = 0.05.$$

Ya que bajo H_0, $\bar{X}$ tiene una distribución normal con media μ_0 y desviación estándar $\sigma/\sqrt{n}$, entonces

$$P\left(Z \geqslant \frac{\bar{x}_0 - \mu_0}{\sigma/\sqrt{n}} \,\middle|\, \mu = \mu_0\right) = 0.05;$$

pero

$$P(Z \geqslant 1.645 \mid \mu = \mu_0) = 0.05,$$

en donde $Z \sim N(0, 1)$. De acuerdo con lo anterior, el valor crítico de $\bar{x}_0$ es tal que:

$$\frac{\bar{x}_0 - \mu_0}{\sigma/\sqrt{n}} = 1.645,$$

o

$$\bar{x}_0 = \frac{1.645\sigma}{\sqrt{n}} + \mu_0.$$

Por lo tanto, se rechazará a H_0: $\mu = \mu_0$ en favor de H_1: $\mu = \mu_1 > \mu_0$ cada vez que un valor de $\bar{X}$ sea $\geqslant (1.645\sigma/\sqrt{n}) + \mu_0$.

Es importante que el lector note que la forma de la mejor región crítica, como está dada por (9.5), para probar H_0: $\mu = \mu_0$ contra H_1: $\mu = \mu_1$ es independiente del valor de μ_1 siempre que $\mu_1 > \mu_0$. En otras palabras, para toda $\mu_1 > \mu_0$ la mejor región crítica en la prueba de H_0: $\mu = \mu_0$ es el extremo derecho de la distribución de muestreo de $\bar{X}$. Así, la expresión (9.5) en realidad da la forma de la mejor región para probar la hipótesis nula simple H_0: $\mu = \mu_0$ contra la hipótesis alternativa compuesta H_1: $\mu > \mu_0$. Esta mejor región crítica recibe el nombre de *región* (o prueba) *uniformemente más potente* para probar H_0: $\mu = \mu_0$ contra H_1: $\mu > \mu_0$. Los comentarios anteriores serán generalizados con la siguiente definición de la mejor prueba.

Definición 9.5 Se dice que una prueba de la hipótesis H_0: $\theta = \theta_0$ es la prueba uniformemente más potente de tamaño α si ésta es por lo menos tan poderosa, para cualquier valor posible θ de la hipótesis alternativa, como cualquier otra prueba de tamaño $\leqslant\alpha$. Esto es, la función de potencia de esta prueba es, por lo menos, tan grande como la de cualquier otra prueba de tamaño $\leqslant\alpha$ para cualquier valor θ de la hipótesis alternativa.

En forma desafortunada no siempre existen las pruebas uniformemente más potentes. Como se ilustró en el ejemplo 9.2, se puede usar el lema de Neyman-Pearson para determinar la prueba uniformemente más potente para cierto número de situaciones de interés práctico en las que la hipótesis alternativa es compuesta pero unilateral.

Ejemplo 9.3 Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño n de una distribución gama con parámetro de escala θ desconocidos y parámetro de forma a.* Determinar la mejor región crítica de tamaño α para probar

$$H_0: \theta = \theta_0$$

contra

$$H_1: \theta = \theta_1,$$

en donde $\theta_1 < \theta_0$.

Se procederá en forma similar a la del ejemplo 9.2. Bajo H_0, la función de verosimilitud es

$$L_0(x_1, x_2, \ldots, x_n; \theta_0) = [\Gamma(a)\theta_0^a]^{-n} \prod_{i=1}^{n} x_i^a \exp\left(-\sum_{i=1}^{n} x_i \Big/ \theta_0\right),$$

y para la hipótesis alternativa ésta es

$$L_1(x_1, x_2, \ldots, x_n; \theta_1) = [\Gamma(a)\theta_1^a]^{-n} \prod_{i=1}^{n} x_i^a \exp\left(-\sum_{i=1}^{n} x_i \Big/ \theta_1\right).$$

Con base en el lema de Neyman-Pearson, la mejor región crítica es aquélla para la cual

$$\frac{\theta_1^{na} \exp\left(-\sum x_i \Big/ \theta_0\right)}{\theta_0^{na} \exp\left(-\sum x_i \Big/ \theta_1\right)} \leqslant k.$$

Esto es

$$\exp\left(-\frac{\sum x_i}{\theta_0} + \frac{\sum x_i}{\theta_1}\right) \leqslant (\theta_0/\theta_1)^{na} k$$

$$\exp\left[\left(\frac{1}{\theta_1} - \frac{1}{\theta_0}\right)\sum x_i\right] \leqslant (\theta_0/\theta_1)^{na} k$$

$$[(\theta_0 - \theta_1)/\theta_0\theta_1]\sum x_i \leqslant \ln[k(\theta_0/\theta_1)^{na}].$$

* Se ha optado por denotar el parámetro de forma de la distribución gama con a en lugar de α para evitar confundir éste con el tamaño del error de tipo I.

Se observa que la cantidad $\theta_0 - \theta_1$ es positiva ya que por hipótesis $\theta_1 < \theta_0$; entonces

$$\sum x_i \leq \frac{\theta_0\theta_1 \ln[k(\theta_0/\theta_1)^{n\alpha}]}{\theta_0 - \theta_1}$$

o

$$\bar{x} \leq \frac{\theta_0\theta_1 \ln[k(\theta_0/\theta_1)^{n\alpha}]}{n(\theta_0 - \theta_1)}. \tag{9.6}$$

De acuerdo con lo anterior, la mejor región crítica para probar H_0: $\theta = \theta_0$ contra H_1: $\theta = \theta_1$ en donde $\theta_1 < \theta_0$ es el extremo izquierdo de la distribución de muestreo de $\bar{X}$. El valor crítico $\bar{x}_0$, para un tamaño dado del error de tipo I, es tal que:

$$P(\bar{X} \leq \bar{x}_0 \mid \theta = \theta_0) = \alpha,$$

y puede encontrarse, en forma directa, de la distribución de $\bar{X}$, la que en este caso también es una distribución gama. Para hacer lo anterior es necesario utilizar la función gama incompleta. De manera alternativa, si el tamaño de la muestra es lo suficientemente grande, puede emplearse el teorema central del límite y usar entonces la aproximación normal.

De nuevo, es interesante notar que la forma de la mejor región crítica dada por (9.6) no depende del valor particular θ siempre que $\theta_1 < \theta_0$. Por tanto, en realidad la región crítica indicada por (9.6) es una región uniformemente más potente para probar H_0: $\theta = \theta_0$ contra H_1: $\theta < \theta_0$ cuando se muestrea una distribución gama con parámetro de forma conocido.

Se invita al lector a que compruebe que si, en el ejemplo 9.2, la hipótesis alternativa es de la forma H_1: $\mu < \mu_0$, la mejor región crítica para probar H_0: $\mu = \mu_0$ es el extremo izquierdo de la distribución de $\bar{X}$. Por lo tanto, se desprende que si en el ejemplo 9.3 la hipótesis alternativa fuese H_1: $\theta > \theta_0$, la mejor región crítica debe ser el extremo derecho de la distribución de $\bar{X}$. Sin embargo, si la hipótesis alternativa en cualquiera de estos dos ejemplos fuese bilateral (esto es, de la forma general H_0: $\theta = \theta_0$ contra H_1: $\theta \neq \theta_0$), no puede encontrarse ninguna región crítica mejor, debido a que para todos los valores alternativos $\theta_1 < \theta_0$, el extremo izquierdo de la distribución de $\bar{X}$ será el mejor, mientras que para todos los valores $\theta_1 > \theta_0$ es el extremo derecho el que será el mejor. Por lo tanto, como regla general, las pruebas uniformemente más potentes usualmente existen para hipótesis alternativas unilaterales, pero éstas no pueden encontrarse para hipótesis alternativas bilaterales.

A continuación se ilustrará el uso del lema de Neyman-Pearson para determinar la mejor región crítica cuando la variable aleatoria de interés es discreta.

Ejemplo 9.4 Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño n de una distribución de Poisson con párametro λ desconocido. Determinar la mejor región crítica de tamaño α para probar

$$H_0: \lambda = \lambda_0$$

contra

$$H_1: \lambda = \lambda_1,$$

donde $\lambda_1 > \lambda_0$.

Al proceder de manera similar a la de los ejemplos 9.2 y 9.3, se tiene

$$L_0(x_1, x_2, \ldots, x_n; \lambda_0) = \frac{\exp(-n\lambda_0)\, \lambda_0^{\Sigma x_i}}{\prod x_i!}$$

y

$$L_1(x_1, x_2, \ldots, x_n; \lambda_1) = \frac{\exp(-n\lambda_1)\, \lambda_1^{\Sigma x_i}}{\prod x_i!}.$$

De esta manera, la mejor región crítica es aquélla para la cual

$$\frac{\exp(-n\lambda_0)\, \lambda_0^{\Sigma x_i}}{\exp(-n\lambda_1)\, \lambda_1^{\Sigma x_i}} \leq k$$

o

$$\left(\frac{\lambda_0}{\lambda_1}\right)^{\Sigma x_i} \exp[n(\lambda_1 - \lambda_0)] \leq k.$$

Después de tomar los logaritmos, se tiene

$$\ln(\lambda_0/\lambda_1) \sum x_i + n(\lambda_1 - \lambda_0) \leq \ln(k)$$

o

$$\ln(\lambda_0/\lambda_1) \sum x_i \leq \ln(k) - n(\lambda_1 - \lambda_0).$$

Pero si $\lambda_1 > \lambda_0$, entonces $0 < \lambda_0/\lambda_1 < 1$ y el logaritmo natural de un número entre 0 y 1 es negativo. Esto da como resultado que la desigualdad anterior pueda escribirse como

$$\sum x_i \geq \frac{\ln(k) - n(\lambda_1 - \lambda_0)}{\ln(\lambda_0/\lambda_1)}. \tag{9.7}$$

La expresión (9.7) define la forma de la mejor región crítica para probar $H_0: \lambda = \lambda_0$ contra $H_1: \lambda = \lambda_1 > \lambda_0$. En particular, dado que $Y = \Sigma X_i$ también es una variable aleatoria de Poisson (véanse los ejercicios en el capítulo 7), la región crítica de la forma $y = \Sigma x_i \geq y_0$ es equivalente a la desigualdad (9.7), en donde el valor crítico y_0 se escoge de manera tal que

$$P(Y \geq y_0) = \alpha.$$

Debido a que Y es una variable aleatoria discreta, es más difícil determinar el valor crítico de y_0 de manera tal que $P(Y \geq y_0)$ sea exactamente igual al tamaño del error de tipo I previamente seleccionado. Para salvar esta dificultad puede implementarse lo

que se conoce como procedimiento de aleatorización (véase [2]). Desde un punto de vista práctico, simplemente se escoge la región crítica y el valor de y_0, cuya área deberá ser lo más cercana al tamaño del error de tipo I que puede tolerarse.

9.5 Principios generales para probar una H_0 simple contra una H_1 uni o bilateral

En la última sección se desarrolló un criterio con el cual se pueden determinar las mejores pruebas para probar hipótesis estadísticas. Se mencionó que no existen pruebas uniformemente más potentes para hipótesis alternativas bilaterales a pesar de que, en forma usual, existen para hipótesis alternativas unilaterales. En esta sección se desarrollarán criterios generales de prueba para los siguientes tres casos los cuales involucran hipótesis nulas simples y alternativas compuestas.

Caso 1	*Caso 2*	*Caso 3*
$H_0: \theta = \theta_0$	$H_0: \theta = \theta_0$	$H_0: \theta = \theta_0$
$H_1: \theta \neq \theta_0$	$H_1: \theta > \theta_0$	$H_1: \theta < \theta_0$

Dado que para el caso 1 no pueden determinarse pruebas uniformemente más potentes, para tipificar éste se desea comparar las funciones de potencia de dos pruebas para un ejemplo específico.

Ejemplo 9.5 Supóngase que en cierta ciudad sólo hay dos estaciones de televisión: el canal 6 y el canal 10. Se piensa que para las noticias de la tarde el auditorio se encuentra dividido en partes iguales para ambos canales. Una compañía se interesa en probar la afirmación de que la proporción de televidentes para las noticias de la tarde es igual a 0.5 para ambos canales. La compañía no posee ninguna información *a priori* para sugerir una alternativa unilateral por lo que decide probar la hipótesis nula

$$H_0: p = 0.5$$

contra

$$H_1: p \neq 0.5.$$

La compañía encuesta a 18 residentes seleccionados al azar y pregunta qué canal prefieren para ver las noticias de la tarde. El número X indica que el canal 6 es el que se ha seleccionado. Se proponen las siguientes dos pruebas:

Prueba A: Rechazar H_0 si $X \leq 4$ o $X \geq 14$.

Prueba B: Rechazar H_0 si $X \leq 5$ o $X \geq 13$.

Si la compañía piensa tolerar un tamaño máximo de 0.1 para el error de tipo I, determinar la mejor prueba a emplear para decidir entre H_0 y H_1.

La estadística de prueba X es una variable aleatoria binomial con $n = 18$ y, bajo la hipótesis nula, $p = 0.5$. Las regiones críticas para ambas pruebas son intuitivamente razonables ya que se rechazará la hipótesis nula para aquellos valores de X que se encuentren cercanos a 0 o a 18. En otras palabras, si p fuese realmente igual a 0.5, debe esperarse observar un valor de X cercano a 9. Entre más se aleje el valor observado del valor de 9, en cualquier dirección, se tendrá más evidencia para inclinarse a rechazar la hipótesis nula. Esto surge del hecho de que cuando se prueban hipótesis estadísticas, el pensamiento se basa estrictamente en la probabilidad. Por ejemplo, si p fuese igual a 0.5, la probabilidad de que X tome un valor entre 6 y 12 incluyendo a estos valores es

$$P(6 \leq X \leq 12) = 0.9038.$$

Por lo tanto, es poco probable que H_0 sea correcta cuando se realice un valor de X grande o pequeño. De hecho, la probabilidad para observar un valor grande o pequeño de X, dado que H_0 es cierta, es precisamente lo que se entiende por el tamaño del error de tipo I.

Para la prueba A, la probabilidad del error de tipo I es

$$\begin{aligned}\alpha_A &= P(X \leq 4 \mid p = 0.5) + P(X \geq 14 \mid p = 0.5)\\ &= 0.0154 + 0.0154\\ &= 0.0308,\end{aligned}$$

y para la prueba B éste es

$$\alpha_B = P(X \leq 5 \mid p = 0.5) + P(X \geq 13 \mid p = 0.5) = 0.0962.$$

No es excesivo notar que las regiones críticas bilaterales son simétricas para ambas pruebas. Esto es lo mejor desde el punto de vista teórico y el procedimiento más aceptado desde el punto de vista práctico para el manejo de hipótesis alternativas bilaterales. Ya que ambas pruebas tienen valores de α menores al tamaño máximo que puede tolerarse del error de tipo I, se compararán sus funciones de potencia para decidir cuál es la mejor de las dos. En la tabla 9.4 se dan las potencias de las pruebas A y B para distintos valores de p.

TABLA 9.4 Funciones de potencia de las pruebas A y B

	p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Prueba A	$P(X \leq 4)$	0.9718	0.7164	0.3327	0.0942	0.0154	0.0013	$\simeq 0$	$\simeq 0$	$\simeq 0$
	$P(X \geq 14)$	$\simeq 0$	$\simeq 0$	$\simeq 0$	0.0013	0.0154	0.0942	0.3327	0.7164	0.9718
	Potencia	0.9718	0.7164	0.3327	0.0955	0.0308	0.0955	0.3327	0.7164	0.9718
Prueba B	$P(X \leq 5)$	0.9936	0.8671	0.5344	0.2088	0.0481	0.0058	0.0003	$\simeq 0$	$\simeq 0$
	$P(X \geq 13)$	$\simeq 0$	$\simeq 0$	0.0003	0.0058	0.0481	0.2088	0.5344	0.8671	0.9936
	Potencia	0.9936	0.8671	0.5347	0.2146	0.0962	0.2146	0.5347	0.8671	0.9936

De la tabla se observa que para cualquier valor de p, la potencia de la prueba B es mayor que la de la prueba A. De acuerdo con lo anterior, la prueba B es uniformemente más poderosa que la prueba A y es la mejor prueba a utilizar para probar las hipótesis indicadas. En la figura 9.5 se dan las curvas de potencia para las pruebas A y B. Nótese que en ambos casos las curvas de potencia crecen en forma simétrica conforme los valores de p se alejan del valor propuesto para éste bajo H_0. Lo anterior es un comportamiento típico de una función de potencia para hipótesis alternativas bilaterales, siempre que la correspondiente región crítica bilateral sea simétrica.

9.5.1 Principios generales para el caso 1

Considérese la prueba de la hipótesis nula

$$H_0: \theta = \theta_0$$

contra la alternativa

$$H_1: \theta \neq \theta_0,$$

donde θ_0 es el valor propuesto de algún parámetro θ bajo H_0. Dada una muestra aleatoria de tamaño n de la distribución de interés, el procedimiento general para probar H_0, es escoger el mejor estimador de θ, T y rechazar H_0 cuando el estimado

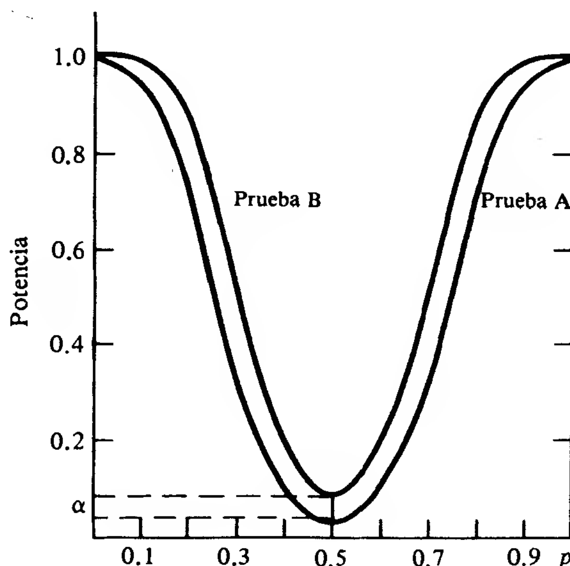


FIGURA 9.5 Comparación de las funciones de potencia para A y B

t obtenido de la muestra, es en forma "suficiente", diferente del valor propuesto de θ_0. Este procedimiento se basa en la noción de un evento raro, la cual ya se ha ilustrado en capítulos anteriores. Esto es, si el estimado t es lo suficientemente distinto del valor propuesto θ_0, entonces se ha observado un evento raro (y la hipótesis nula es correcta), o se ha observado un valor de la estadística que sugiere un valor θ diferente del propuesto θ_0. Cuando el estimado t es en forma suficiente distinto de θ_0, se asumirá la última posibilidad y se dejará el tamaño del error de tipo I igual a la probabilidad del anterior. En particular, para un tamaño preseleccionado α, del error de tipo I se obtiene una región crítica bilateral en los extremos de la distribución de muestreo de T, de manera tal que el área, en cualquier lado, más allá del valor crítico es igual a $\alpha/2$. Entonces se rechaza H_0 en favor de H_1 cuando el estimado t se encuentra dentro de la región crítica. Cuando el estimado t no se encuentra dentro de la región crítica, no puede rechazarse la hipótesis nula. De esta forma, cualquier diferencia con respecto al valor de θ_0 se considera causada por la fluctuación en el muestreo del estimador T.

Este enfoque es muy similar a la construcción de un intervalo de confianza bilateral para θ. Para cualquier valor propuesto de θ_0 que se encuentre dentro de un intervalo de confianza del $100(1 - \alpha)\%$ para θ, H_0 no será rechazada. Dado un intervalo de confianza del $100(1 - \alpha)\%$ para θ, sólo los valores propuestos bajo H_0 que se encuentren fuera de este intervalo darán como resultado el rechazo de la hipótesis nula. En este contexto, es apropiado considerar a un intervalo de confianza como una proposición más general de inferencia estadística para θ, ya que ésta incluye a todos los posibles valores de θ_0 que podrían no llevar al rechazo de la hipótesis nula.

9.5.2. Principios generales para el caso 2

Considérese la hipótesis nula

$$H_0: \theta = \theta_0$$

contra la alternativa

$$H_1: \theta > \theta_0.$$

Para este caso al igual que para el caso tres, la naturaleza unilateral de la hipótesis alternativa sugiere la existencia de alguna información *a priori* la cual ayuda a definir la dirección unilateral de H_1 en relación con el valor propuesto de θ_0. El procedimiento general para probar H_0 es de nuevo el escoger la mejor estadística T de θ y rechazar H_0 cuando el estimado t es en forma "suficiente" mayor que el valor propuesto θ_0. La palabra "suficiente" implica que se tiene una tolerancia para la fluctuación en el muestreo del estimador T. Sin embargo, si lo que se obtiene de esta forma por medio de la muestra aleatoria se encuentra más allá de esta tolerancia, H_0 será rechazada. De esta forma, para un tamaño α, del error de tipo I, la región crítica se encuentra localizada en el extremo superior de la distribución de muestreo de T y H_0 se rechaza si el estimado t no es menor que el valor crítico. En la figura 9.6 se ilustra la curva de potencia típica para este caso.

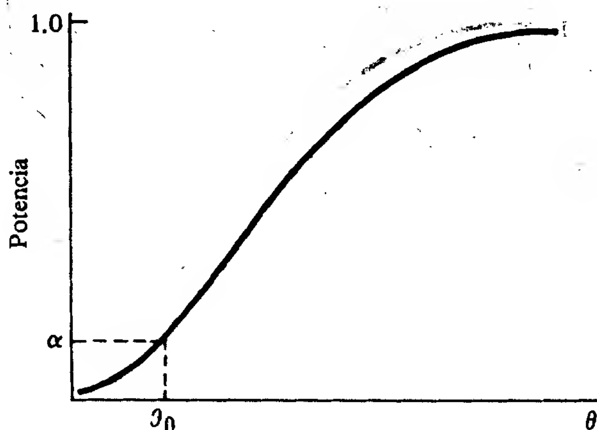


FIGURA 9.6 Curva típica de potencia para el caso 2

9.5.3 Principios generales para el caso 3

Para probar la hipótesis

$$H_0: \theta = \theta_0$$

contra

$$H_1: \theta < \theta_0,$$

el procedimiento general es rechazar a H_0 cada vez que el estimado t sea, en forma "suficiente", menor que el valor propuesto θ_0. La región crítica de tamaño α se localiza en el extremo inferior de la distribución de muestreo de T en forma tal que el área a la izquierda del valor crítico sea igual al tamaño α del error de tipo I. Cualquier valor t de la estadística de prueba T que se encuentre en la región crítica llevará al rechazo de H_0. En la figura 9.7 se muestra la curva de potencia para este caso.

Con respecto a la prueba de hipótesis estadísticas, el lector debe tomar nota de lo siguiente. Debido a que se coloca un gran enfásis en el tamaño del error de tipo I generalmente se formula la hipótesis nula en forma tal que ésta se rechace si la evidencia experimental apoya esta decisión. En otras palabras, lo que realmente se desea es concluir que la hipótesis alternativa es la correcta. De esta forma, cuando se prueban hipótesis estadísticas, se juega un papel parecido al de un fiscal en su intento de proporcionar la suficiente evidencia para rechazar la hipótesis nula. Los indicado es escoger el tamaño del error de tipo I antes de la determinación de la muestra aleatoria. Si se obtiene como resultado que la hipótesis nula no puede rechazarse con el valor escogido de α debe evitarse aumentar el tamaño del error de tipo I con la idea de rechazar la hipótesis nula.

La discusión anterior constituye el método clásico para probar hipótesis estadísticas. Se han dirigido algunas críticas directas hacia este enfoque debido a que la de-

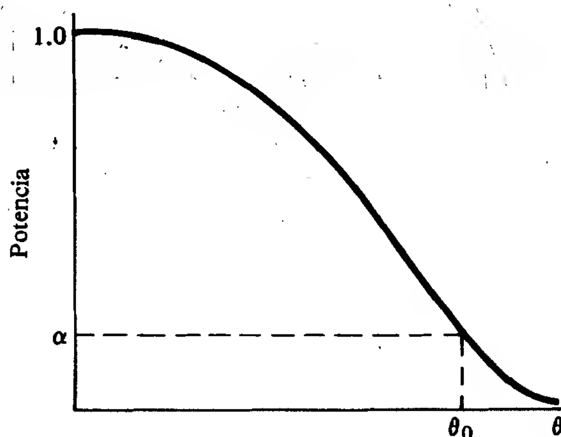


FIGURA 9.7 Curva típica de potencia para el caso 3

cisión final de rechazar o no una H_0 dada, es demasiado cortante y seca y no proporciona una medida real de que la decisión sea correcta en términos de la probabilidad. Para esto lo que se ha sugerido es el cálculo del llamado valor *p*. El valor *p* es la probabilidadd, dado que H_0 es cierta, de que la estadística de prueba tome un valor mayor o igual que el calculado con base en la muestra aleatoria. Un valor *p* relativamente pequeño puede sugerir que si H_0 es realmente cierta, el valor de la estadística de prueba sea poco probable. Puede entonces optarse por rechazar H_0 debido a que esta decisión tendrá una alta probabilidad de ser correcta.

Se recomienda el cálculo del valor *p* acoplado con el enfoque clásico de escoger un tamaño del error de tipo I antes de la determinación de la muestra aleatoria. Entonces, la decisión de rechazar o no a H_0 puede basarse en una región crítica de tamaño α, con el valor *p* proporcionando una medida real en términos de la probabilidad de que la decisión sea correcta. De acuerdo con lo anterior, se sugiere la siguiente regla: si el valor *p* es menor o igual a α, se rechaza H_0; de otra forma no puede rechazarse la hipótesis nula. El cálculo del valor *p* se ilustrará en los ejemplos subsecuentes de este capítulo. Debe notarse que muchos paquetes estadísticos para computadora, tales como SAS, SPSS, BMD y otros, imprimen el valor *p* para casi todas las situaciones en las que se involucra, de alguna manera, la prueba de hipótesis estadísticas.

9.6 Prueba de hipótesis con respecto a las medias cuando se muestrean distribuciones normales

En esta sección se estudiará la prueba de hipótesis sobre la media de una distribución normal o las medias de dos distribuciones normales independientes. Se examinarán

los casos en los que los valores de las varianzas son tanto conocidos como no conocidos. Se invita al lector a que consulte las secciones 8.4.1 a 8.4.3 para efectuar comparaciones con los intervalos de confianza.

9.6.1 Pruebas para una muestra

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución normal con media μ desconocida. En este caso el interés recae en probar uno de los siguientes conjuntos de hipótesis con respecto a μ.

$$H_0: \mu = \mu_0 \qquad H_0: \mu = \mu_0 \qquad H_0: \mu = \mu_0$$
$$H_1: \mu \neq \mu_0 \qquad H_1: \mu > \mu_0 \qquad H_1: \mu < \mu_0$$

Primero, supóngase que el valor de la varianza poblacional σ^2 es conocido. Entonces la estadística de prueba es la media muestral $\overline{X}$, misma que, bajo la hipótesis nula, tiene una distribución normal con media μ_0 y desviación estándar $\sigma/\sqrt{n}$. La región crítica de tamaño α para la hipótesis bilateral es de la forma

$$\text{Rechazar } H_0 \text{ si } \begin{cases} \overline{X} \geqslant \bar{x}_{1-\alpha/2} \\ \quad \text{or} \\ \overline{X} \leqslant \bar{x}_{\alpha/2}, \end{cases} \tag{9.8}$$

donde $\bar{x}_{1-\alpha/2}$ y $\bar{x}_{\alpha/2}$ son los valores cuantiles críticos de $\overline{X}$ de manera tal que

$$P(\overline{X} \geqslant \bar{x}_{1-\alpha/2}) = \alpha/2 \quad \text{and} \quad P(\overline{X} \leqslant \bar{x}_{\alpha/2}) = \alpha/2.$$

Dado que bajo H_0, $\overline{X} \sim N(\mu_0, \sigma/\sqrt{n})$, entonces en forma equivalente

$$P\left(Z \geqslant \frac{\bar{x}_{1-\alpha/2} - \mu_0}{\sigma/\sqrt{n}}\right) = \alpha/2 \quad \text{y} \quad P\left(Z \leqslant \frac{\bar{x}_{\alpha/2} - \mu_0}{\sigma/\sqrt{n}}\right) = \alpha/2$$

o

$$z_{1-\alpha/2} = \frac{\bar{x}_{1-\alpha/2} - \mu_0}{\sigma/\sqrt{n}} \quad \text{y} \quad z_{\alpha/2} = \frac{\bar{x}_{\alpha/2} - \mu_0}{\sigma/\sqrt{n}},$$

en donde $z_{1-\alpha/2}$ y $z_{\alpha/2}$ son los correspondientes valores cuantiles de Z. Por lo tanto, se sigue que H_0 debe rechazarse cuando un valor $\bar{x}$ de la media muestral $\overline{X}$ es tal que

$$\bar{x} \geqslant \frac{\sigma\, z_{1-\alpha/2}}{\sqrt{n}} + \mu_0 \quad \text{o} \quad \bar{x} \leqslant \frac{\sigma\, z_{\alpha/2}}{\sqrt{n}} + \mu_0.$$

De manera equivalente, se rechazará H_0 cuando

$$z \geqslant z_{1-\alpha/2} \quad \text{o} \quad z \leqslant z_{\alpha/2},$$

donde $z = (\bar{x} - \mu_0)/(\sigma/\sqrt{n})$ es el valor de la correspondiente normal estándar al valor $\bar{x}$ de $\overline{X}$.

Para la hipótesis alternativa unilateral H_1: $\mu > \mu_0$, la región crítica de tamaño α es el extremo derecho de la distribución de muestreo de $\overline{X}$; ésta es de la forma

$$\text{Rechazar } H_0 \quad \text{si } \overline{X} \geq \bar{x}_{1-\alpha}, \tag{9.9}$$

en donde $\bar{x}_{1-\alpha}$ es el valor cuantil de $\overline{X}$, tal que $P(\overline{X} \geq \bar{x}_{1-\alpha}) = \alpha$. En forma similar, para la hipótesis alternativa H_1: $\mu < \mu_0$, la región crítica es de la forma

$$\text{Rechazar } H_0 \quad \text{si } \overline{X} \leq \bar{x}_{\alpha}, \tag{9.10}$$

en donde el valor $\bar{x}_\alpha$ es tal que $P(\overline{X} \leq \bar{x}_\alpha) = \alpha$.

En la figura 9.8 se ilustran las regiones críticas para las hipótesis unilaterales en términos de la estadística $\overline{X}$ y su transformación a la variable aleatoria normal estándar Z. En la tabla 9.5 se proporciona un resumen de los criterios de rechazo para la pueba de hipótesis con respecto a la media de una distribución normal con varianza conocida.

Antes de resolver un ejemplo, se desarrollará una expresión general para la determinación del error de tipo II para uno de los casos. Considérese la hipótesis nula H_0: $\mu = \mu_0$ contra la alternativa H_1: $\mu > \mu_0$. Supóngase que en realidad $\mu = \mu_1 > \mu_0$. De acuerdo con (9.9), no puede rechazarse H_0 si un valor de $\overline{X}$ es menor que $(\sigma z_{1-\alpha}/\sqrt{n}) + \mu_0$. Dado que la probabilidad del error de tipo II es igual

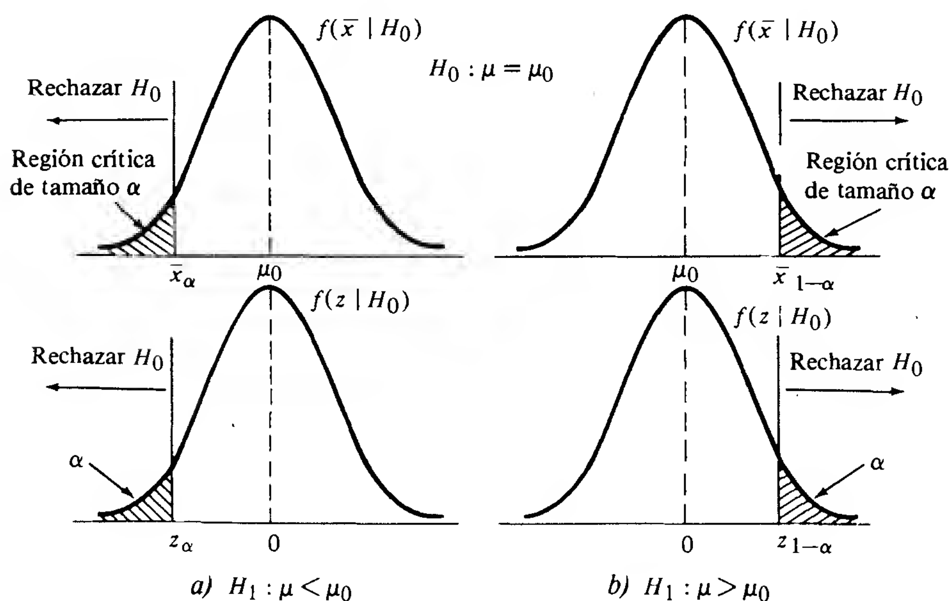


FIGURA 9.8 Regiones críticas para hipótesis alternativas unilaterales

TABLA 9.5 Criterios de rechazo para la prueba de hipótesis con respecto a la media de una distribución normal con varianza conocida

Hipótesis nula	*Valor de la estadística de prueba bajo H_0*
H_0: $\mu = \mu_0$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
Hipótesis alternativa	*Criterios de rechazo*
H_1: $\mu \neq \mu_0$	Rechazar H_0 cuando $z \leq z_{\alpha/2}$ o cuando $z \geq z_{1-\alpha/2}$
H_1: $\mu > \mu_0$	Rechazar H_0 cuando $z \geq z_{1-\alpha}$
H_1: $\mu < \mu_0$	Rechazar H_0 cuando $z \leq z_\alpha$

a la probabilidad de no rechazar un H_0 falsa, es necesario determinar

$$\beta = P\left(\bar{X} < \frac{\sigma z_{1-\alpha}}{\sqrt{n}} + \mu_0 \,\middle|\, \mu = \mu_1 > \mu_0\right),$$

la que en términos de la normal estándar es

$$\beta = P\left[Z < \frac{\dfrac{\sigma z_{1-\alpha}}{\sqrt{n}} + \mu_0 - \mu_1}{\sigma/\sqrt{n}} \,\middle|\, \mu = \mu_1\right]. \tag{9.11}$$

Al sustituir cualquier valor μ_1 de la hipótesis alternativa en (9.11), se puede calcular el correspondiente valor de la probabilidad del error de tipo II y, de esta forma, la potencia. Nótese que β (y la potencia) dependen del tamaño de la muestra n, del tamaño α, del error de tipo I, de la diferencia $(\mu_0 - \mu_1)$ entre el valor propuesto μ_0 bajo H_0 y el verdadero valor μ_1 bajo H_1, y de la desviación estándar σ de la población. Para un valor fijo de α, $(\mu_0 - \mu_1)$ y σ, el tamaño del error de tipo II disminuye conforme n aumenta. Para valores fijos de n, $(\mu_0 - \mu_1)$ y σ, β aumenta conforme α disminuye. Y para valores fijos de n, α, y σ, β disminuye conforme la diferencia $(\mu_0 - \mu_1)$ aumenta.

Para otros casos, se pueden desarrollar expresiones similares a (9.11). El comportamiento general del tamaño del error de tipo II como una función de n, α, $(\mu_0 - \mu_1)$, y σ es igual al anterior.

Ejemplo 9.6 Los siguientes datos representan los tiempos de armado para 20 unidades seleccionadas aleatoriamente: 9.8, 10.4, 10.6, 9.6, 9.7, 9.9, 10.9, 11.1, 9.6, 10.2, 10.3, 9.6, 9.9, 11.2, 10.6, 9.8, 10.5, 10.1, 10.5, 9.7. Supóngase que el tiempo necesario para armar una unidad es una variable aleatoria normal con media μ y desviación etándar $\sigma = 0.6$ minutos. Con base en esta muestra, ¿existe alguna razón para creer, a un nivel de 0.05, que el tiempo de armado promedio es mayor de 10 minutos?

Considérese la hipótesis nula

$$H_0: \mu = 10$$

contra la alternativa

$$H_1: \mu > 10.$$

Si puede rechazarse a H_0 con $\alpha = 0.05$, entonces existe una razón para creer que el tiempo necesario para armar una unidad es mayor de 10 minutos. Dado que $P(Z \geq 1.645) = 0.05$, el valor crítico en términos de la variable aleatoria normal estándar es $z_{0.95} = 1.645$. De los datos de la muestra, el valor $\bar{x}$ es igual a 10.2 minutos. Entonces

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{10.2 - 10}{0.6/\sqrt{20}} = 1.4907.$$

Dado que $z = 1.4907 < z_{0.95} = 1.645$, no puede rechazarse la hipótesis nula. El valor p en este caso es la probabilidad de que la variable aleatoria normal estándar sea mayor o igual al valor de 1.4907, dando como resultado que H_0 sea cierta. Puede verse, de la tabla D del apéndice, que

$$P(Z \geq 1.4907 \mid \mu = 10) = 0.0681.$$

Puesto que $p = 0.0681 > \alpha = 0.05$, se concluye que con base en la muestra no existe la suficiente evidencia para rechazar la hipótesis de que el tiempo promedio necesario para armar una unidad es de 10 minutos.

En el contexto de este ejemplo, supóngase que se desea dar respuesta a la siguiente pregunta. Si el verdadero tiempo promedio necesario para armar una unidad es de 10.3 minutos, ¿cuál es la probabilidad de rechazar la hipótesis nula? En este caso se desea obtener la potencia de la prueba para detectar la falta de veracidad de H_0 cuando el valor verdadero es de 10.3 minutos. Primero se obtendrá el tamaño del error de tipo II. Mediante el uso de (9.11) se tiene

$$\begin{aligned}\beta &= P\left(Z < \frac{\frac{(0.6)(1.645)}{\sqrt{20}} + 10 - 10.3}{0.6/\sqrt{20}} \,\middle|\, \mu = 10.3\right)\\ &= P(Z < -0.59 \mid \mu = 10.3)\\ &= 0.2776.\end{aligned}$$

De esta forma la probabilidad de equivocarse al rechazar H_0 cuando la media es 10.3 minutos, es igual a 0.2776. Por lo tanto, potencia = $1 - \beta = 0.7224$. Si se sigue este procedimiento se obtienen β y las probabilidades de potencia para otros valores de μ bajo la hipótesis alternativa, tal y como se encuentran resumidos en la tabla 9.6. Nótese que conforme la diferencia entre el valor propuesto de la media bajo H_0 y el valor verdadero bajo H_1 aumenta, la potencia de la prueba también aumenta.

Supóngase que se tiene la misma situación pero con la excepción de que no se conoce el valor de la varianza poblacional σ^2. Con base en la sección 8.4.2, la mejor es-

TABLA 9.6 Error de tipo II y probabilidades de potencia para el ejemplo 9.6

μ	*10.01*	*10.1*	*10.2*	*10.3*	*10.4*	*10.5*	*10.6*	*10.7*
β	0.9418	0.8159	0.5596	0.2776	0.0901	0.0188	0.0024	0.0002
Potencia	0.0582	0.1841	0.4404	0.7224	0.9099	0.9812	0.9976	0.9998

tadística de prueba a utilizar en este caso tiene una distribución t de Student. Éste es, bajo la hipótesis nula H_0: $\mu = \mu_0$ la estadística

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

tiene una distribución t de Student con $n - 1$ grados de libertad. El lector debe tener muy poca dificultad al reconocer que mediante el empleo de la distribución t de Student, las regiones críticas para este caso son similares a las del caso anterior con respecto a las hipótesis alternativas uni o bilaterales. En la tabla 9.7 se proporciona un resumen.

Ejemplo 9.7 Mediante el empleo de los datos del ejemplo 8.9, demostrar que para cualquier valor propuesto μ_0 para μ que se encuentre en el interior de un intervalo de confianza del 95%, una prueba de la hipótesis

$$H_0: \mu = \mu_0$$

contra la alternativa

$$H_1: \mu \neq \mu_0$$

no llevará al rechazo de H_0 para $\alpha = 0.05$.

Recuérdese la sección 8.4.2 en la que un intervalo del 95% de confianza para μ es 500.45–507.05. Es necesario demostrar que los límites 500.45 y 507.05 coinciden

TABLA 9.7 Criterios de rechazo para probar hipótesis con respecto a la media de una distribución normal con varianza desconocida

Hipótesis nula	*Valor de la estadística de prueba bajo* H_0
H_0: $\mu = \mu_0$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$
Hipótesis alternativa	*Criterios de rechazo*
H_1: $\mu \neq \mu_0$	Rechazar H_0 cuando $t \leq t_{\alpha/2,n-1}$ o cuando $t \geq t_{1-\alpha/2,n-1}$
H_1: $\mu > \mu_0$	Rechazar H_0 cuando $t \geq t_{1-\alpha,n-1}$
H_1: $\mu < \mu_0$	Rechazar H_0 cuando $t \leq t_{\alpha,n-1}$

con los límites de los valores propuestos μ_0 bajo H_0 que llevan al rechazo de la hipótesis nula. Dado que $\bar{x} = 503.75$ y $s = 6.2$ para el límite 500.45 se tiene

$$t = \frac{503.75 - 500.45}{6.2/\sqrt{16}} = 2.131,$$

y para el límite 507.05

$$t = \frac{503.75 - 507.05}{6.2/\sqrt{16}} = -2.131.$$

Pero los valores ± 2.131 son los límites de la región crítica bilateral de tamaño $\alpha = 0.05$ y 15 grados de libertad. En otras palabras, si $\mu_0 \leqslant 500.45$, entonces $t \geqslant 2.131$, y si $\mu_0 \geqslant 507.05, t \leqslant -2.131$. De esta forma, cualquier valor propuesto μ_0 interior a 500.45 y 507.05 no llevará al rechazo de H_0 con $\alpha = 0.05$.

Para ilustrar el cálculo del valor p en el contexto de este ejemplo, considérese la siguiente hipótesis nula

$$H_0: \mu = 508$$

contra la alternativa

$$H_1: \mu \neq 508.$$

Dado que el valor propuesto de 508 se encuentra fuera del intervalo de confianza del 95%, H_0 será rechazada a un nivel $\alpha = 0.05$. Para obtener el valor p se calcula el valor de la estadística de prueba, el cual es

$$t = \frac{503.75 - 508}{6.2/\sqrt{16}} = -2.742.$$

Dado que la hipótesis alternativa es bilateral, el valor p está dado por

$$P(|T| \geqslant 2.742) = P(T \leqslant -2.742) + P(T \geqslant 2.742),$$

en donde T es una variable aleatoria t de Student con 15 grados de libertad. En la tabla F del apéndice puede observarse que es necesario interpolar entre los valores cuantiles $t_{0.99,\,15} = 2.602$ y $t_{0.995,\,15} = 2.947$. Entonces $t_{0.992,\,15} = 2.742$, y el valor p es, en forma aproximada, 0.016. Por lo tanto, si la hipótesis nula es cierta, existe una oportunidad menor del 2% para observar un valor de la distribución t de Student con 15 grados de libertad cuya magnitud sea igual o mayor al valor observado de 2.742.

La determinación de la potencia y de las probabilidades de los errores de tipo II para la estadística T es algo más difícil que en el caso previo, el cual involucraba una distribución normal. La dificultad surge debido a que la distribución de la estadística de prueba, si H_0 es falsa, no es exactamente igual a la distribución t de Student. De hecho, bajo la hipótesis alternativa la estadística tiene lo que se conoce como una *distribución t no central,* la cual difiere de la ordinaria t de Student por la introducción de un parámetro adicional. El parámetro, denotado por δ, se define

por

$$\delta = \frac{\mu - \mu_0}{\sigma}$$

y expresa la diferencia entre el verdadero valor de μ bajo H_1 y el valor propuesto μ_0 bajo H_0 en términos de σ. Como resultado se tiene que la función de potencia de la estadística T depende tanto de los grados de libertad ν y de δ. En este caso existen las curvas CO como funciones de δ y del tamaño de la muestra n tanto para las hipótesis alternativas unilaterales como para las bilaterales (véase [1]). Éstas revelan el mismo comportamiento para el tamaño del error de tipo II con respecto a n, α, y la diferencia entre los valores bajo H_1 y H_0 al igual que en el caso previo. Debe notarse que para muestras de tamaño relativamente grande, por ejemplo mayor que 30, el cálculo de la potencia para la estadística T se puede manejar en forma adecuada mediante el empleo de la aproximación normal.

9.6.2 Pruebas para dos muestras

Sean $X_1, X_2, \ldots, X_{n_X}$ y $Y_1, Y_2, \ldots, Y_{n_Y}$ muestras aleatorias provenientes de dos distribuciones normales independientes con medias μ_X y μ_Y y varianzas σ_X^2 y σ_Y^2, respectivamente. Supóngase que se desea probar la hipótesis nula

$$H_0: \mu_X - \mu_Y = \delta_0$$

contra una de las siguientes alternativas:

$$H_1: \mu_X - \mu_Y \neq \delta_0 \qquad H_1: \mu_X - \mu_Y > \delta_0 \qquad H_1: \mu_X - \mu_Y < \delta_0,$$

en donde δ_0 es una cantidad que toma valores positivos o cero y la cual representa la diferencia propuesta entre los valores desconocidos de las medias. Supóngase que las varianzas de la población se conocen. De las discusiones en las secciones 7.7, 8.4.3 y el material precedente de este capítulo, es razonable concluir que la estadística de prueba apropiada es la diferencia muestral media $\overline{X} - \overline{Y}$. En particular, si un valor de $\overline{X} - \overline{Y}$ con base en la muestra aleatoria es lo suficientemente diferente, mayor o menor que δ_0, se rechazará la hipótesis nula dependiendo de la hipótesis alternativa en cuestión. Una transformación a la distribución normal estándar da origen a una forma equivalente de la prueba estadística dada por (8.41). En la tabla 9.8 se proporciona un resumen de la información pertinente para este caso.

Ejemplo 9.8 Supóngase que se tienen muestras aleatorias de igual tamaño n de dos distribuciones normales independientes con varianzas conocidas σ_X^2 y σ_Y^2, las cuales se emplean para probar la hipótesis nula

$$H_0: \mu_X - \mu_Y = \delta_0$$

contra la alternativa

$$H_1: \mu_X - \mu_Y = \delta_1 > \delta_0.$$

TABLA 9.8 Criterios de rechazo para la prueba de hipótesis con respecto a las medias de dos distribuciones normales e independientes con varianzas conocidas

Hipótesis nula	*Valor de la estadística de prueba bajo H_0*
$H_0\colon \mu_X - \mu_Y = \delta_0$	$z = \dfrac{\bar{x} - \bar{y} - \delta_0}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}}$
Hipótesis alternativa	*Criterios de rechazo*
$H_1\colon \mu_X - \mu_Y \neq \delta_0$	Rechazar H_0 cuando $z \leqslant z_{\alpha/2}$ o cuando $z \geqslant z_{1-\alpha/2}$
$H_1\colon \mu_X - \mu_Y > \delta_0$	Rechazar H_0 cuando $z \geqslant z_{1-\alpha}$
$H_1\colon \mu_X - \mu_Y < \delta_0$	Rechazar H_0 cuando $z \leqslant z_\alpha$

Si se especifican los tamaños particulares α y β de los errores de tipo I y de tipo II, respectivamente, obtener una expresión para n.

Si H_0 es realmente cierta, la probabilidad de rechazarla es α; y si H_0 es falsa ($\mu_X - \mu_Y = \delta_1 > \delta_0$), la probabilidad de no rechazar H_0 es β. Sea c_0 el valor crítico con respecto a la distribución de muestreo de $\overline{X} - \overline{Y}$. Entonces H_0 será rechazada cuando $\bar{x} - \bar{y} \geqslant c_0$, tal que

$$P(\overline{X} - \overline{Y} \geqslant c_0 \mid \mu_X - \mu_Y = \delta_0) = \alpha.$$

En términos de la variable aleatoria normal estándar, lo anterior es equivalente a

$$P\left(Z \geqslant \frac{c_0 - \delta_0}{\sqrt{\dfrac{\sigma_X^2 + \sigma_Y^2}{n}}} \,\middle|\, \mu_X - \mu_Y = \delta_0\right) = \alpha.$$

Dado que pueden determinarse valores cuantiles $z_{1-\alpha}$ de la normal entándar tales que

$$P(Z \geqslant z_{1-\alpha}) = \alpha,$$

se tiene

$$\frac{c_0 - \delta_0}{\sqrt{\dfrac{\sigma_X^2 + \sigma_Y^2}{n}}} = z_{1-\alpha}. \tag{9.12}$$

Si $\mu_X - \mu_Y = \delta_1 > \delta_0$, entonces la probabilidad de no rechazar a H_0 es β. Por lo tanto

$$P(\overline{X} - \overline{Y} < c_0 \mid \mu_X - \mu_Y = \delta_1) = \beta,$$

que en términos de la variable normal estándar es

$$P\left(Z < \frac{c_0 - \delta_1}{\sqrt{\dfrac{\sigma_X^2 + \sigma_Y^2}{n}}} \,\middle|\, \delta_1\right) = \beta.$$

Pero el valor cuantil z_β debe ser un punto de la normal estándar tal que

$$P(Z < z_\beta) = \beta.$$

Entonces se sigue que

$$\frac{c_0 - \delta_1}{\sqrt{\dfrac{\sigma_X^2 + \sigma_Y^2}{n}}} = z_\beta. \tag{9.13}$$

Debe notarse que puesto que es poco probable que β sea menor que 0.05, el valor cuantil z_β es negativo.

Nótese que las ecuaciones (9.12) y (9.13) contienen dos incógnitas: c_0 y n. Para resolver para n, primero se resolverán ambas ecuaciones para c_0.

$$c_0 = z_{1-\alpha}\sqrt{\frac{\sigma_X^2 + \sigma_Y^2}{n}} + \delta_0,$$

y

$$c_0 = z_\beta\sqrt{\frac{\sigma_X^2 + \sigma_Y^2}{n}} + \delta_1.$$

Al igualar ambos miembros derechos, se tiene

$$z_{1-\alpha}\sqrt{\frac{\sigma_X^2 + \sigma_Y^2}{n}} + \delta_0 = z_\beta\sqrt{\frac{\sigma_X^2 + \sigma_Y^2}{n}} + \delta_1,$$

o

$$\sqrt{\frac{\sigma_X^2 + \sigma_Y^2}{n}}\,(z_{1-\alpha} - z_\beta) = \delta_1 - \delta_0.$$

Dado que para la normal estándar $-z_\beta = z_{1-\beta}$,

$$\sqrt{\frac{\sigma_X^2 + \sigma_Y^2}{n}}\,(z_{1-\alpha} + z_{1-\beta}) = \delta_1 - \delta_0.$$

la cual, después de resolver para n, se reduce a

$$n = \frac{(\sigma_X^2 + \sigma_Y^2)\,(z_{1-\alpha} + z_{1-\beta})^2}{(\delta_1 - \delta_0)^2}. \tag{9.14}$$

La expresión (9.14) determina el tamaño de cada una de las dos muestras aleato-

rias en las dos distribuciones normales independientes, asegurando probabilidades α y β para los errores de tipo I y tipo II, respectivamente, cuando se prueba

$$H_0: \mu_X - \mu_Y = \delta_0$$

contra

$$H_1: \mu_X - \mu_Y = \delta_1 > \delta_0.$$

Para un ejemplo específico, sean $\sigma_X^2 = 25$, $\sigma_Y^2 = 20$, $\delta_0 = 5$, $\delta_1 = 8$, $\alpha = 0.05$, y $\beta = 0.10$. Entonces $z_{0.95} = 1.645$, $z_{0.90} = 1.28$, y

$$n = \frac{(25 + 20)(1.645 + 1.28)^2}{(8 - 5)^2} = 43.$$

Se invita al lector a que obtenga una expresión similar para la hipótesis alternativa del lado izquierdo. Para una hipótesis alternativa bilateral, es posible obtener una aproximación del tamaño de n mediante el empleo de la expresión (9.14) y reemplazando α con $\alpha/2$. A pesar de que este enfoque no es exacto, para muchas situaciones prácticas es suficiente.

A continuación se examinará el caso en el que el valor de la varianza no se conoce; si las varianzas σ_X^2 y σ_Y^2 no se conocen pero se supone que son iguales, entonces para la hipótesis nula

$$H_0: \mu_X - \mu_Y = \delta_0$$

la estadística de prueba es

$$T = \frac{\overline{X} - \overline{Y} - \delta_0}{S_p\sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}}, \qquad (9.15)$$

la cual tiene una distribución t de Student con $n_X + n_Y - 2$ grados de libertad. El estimador combinado S_p^2 de la varianza común σ^2 está dado por la expresión (7.28). De las discusiones anteriores, las regiones críticas de tamaño α para las hipótesis alternativas uni y bilateral, deben ser evidentes. Éstas se encuentran resumidas en la tabla 9.9.

Ejemplo 9.9 En forma reciente se ha incrementado el interés de evaluar el efecto del ruido sobre la habilidad de las personas para llevar a cabo una determinada tarea. Un investigador diseña un experimento en el que se pedirá a un determinado número de sujetos que lleven a cabo una tarea específica en un medio controlado y bajo dos niveles diferentes de ruido de fondo. El investigador selecciona 32 personas que son capaces de realizar la misma tarea y de manera práctica en el mismo tiempo. Del total de personas, 16 seleccionadas al azar realizarán esta tarea bajo un nivel modesto de ruido de fondo. Las restantes 16 llevarán a cabo la misma tarea bajo un ruido de nivel 2, el cual es más severo que el ruido de nivel 1. Los siguientes datos representan los tiempos observados (en minutos) que fueron necesarios para completar la tarea para cada una de las 16 personas de cada nivel.

TABLA 9.9 Criterios de rechazo para la prueba de hipótesis con respecto a las medias de dos distribuciones normales e independientes con varianzas iguales pero desconocidas

Hipótesis nula	*Valor de la estadística de prueba bajo H_0*
$H_0: \mu_X - \mu_Y = \delta_0$	$t = \dfrac{\bar{x} - \bar{y} - \delta_0}{s_p\sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}}$
Hipótesis alternativa	***Criterios de rechazo***
$H_1: \mu_X - \mu_Y \neq \delta_0$	Rechazar H_0 cuando $t \leq t_{\alpha/2,m}$ o cuando $t \geq t_{1-\alpha/2,m}$, en donde $m = n_X + n_Y - 2$
$H_1: \mu_X - \mu_Y > \delta_0$	Rechazar H_0 cuando $t \geq t_{1-\alpha,m}$
$H_1: \mu_X - \mu_Y < \delta_0$	Rechazar H_0 cuando $t \leq t_{\alpha,m}$

Nivel 1	14	12	15	15	11	16	17	12	14	13	18	13	18	15	16	11
Nivel 2	20	22	18	18	19	15	18	15	22	18	19	15	21	22	18	16

Asumiendo que estos datos constituyen muestras aleatorias de dos distribuciones normales e independientes con varianzas iguales pero no conocidas, ¿existe alguna razón para creer que el tiempo promedio para el nivel 2 es mayor por más de dos minutos que para el nivel 1 con $\alpha = 0.01$?

Sean μ_1 y μ_2 las medias desconocidas para los niveles 1 y 2 respectivamente. El valor propuesto para la diferencia entre μ_2 y μ_1 es $\delta_0 = 2$. En otras palabras, se afirma que el valor de μ_2 es mayor que μ_1 por una cantidad igual a dos minutos; pero en realidad lo que se desea demostrar es que μ_2 es más grande que μ_1 por más de dos minutos. De acuerdo con lo anterior, considérese la hipótesis nula

$$H_0: \mu_2 - \mu_1 = 2$$

contra la alternativa

$$H_1: \mu_2 - \mu_1 > 2.$$

Dado que $\alpha = 0.01$ y $n_1 = n_2 = 16$, el valor crítico es $t_{0.99,30} = 2.457$. De los datos se tiene que $\bar{x}_1 = 14.375$, $\bar{x}_2 = 18.5$, $s_1 = 2.2767$, y $s_2 = 2.4495$; por lo que el estimado combinado de la varianza común es

$$s_p^2 = \frac{(15)(2.2767)^2 + 15(2.4495)^2}{16 + 16 - 2} = 5.5917,$$

o

$$s_p = 2.3647.$$

Entonces el valor de la estadística de prueba es

$$t = \frac{(18.5 - 14.375) - 2}{2.3647\sqrt{\dfrac{1}{16} + \dfrac{1}{16}}} = 2.5417.$$

Dado que el valor de 2.5417 se encuentra dentro de la región crítica de tamaño 0.01, se rechaza la hipótesis nula. Bajo H_0, el valor p es la probabilidad de que $T \geq 2.5417$, en donde $T \sim t$ de Student con 30 grados de libertad. Mediante el empleo de la tabla F del apéndice y después de interpolar, se obtiene que

$$P(T \geq 2.5417) = 0.0085.$$

Por lo tanto, con base en este experimento, puede concluirse que la diferencia entre las medias de los niveles 1 y 2 es mayor de dos minutos estadísticamente discernible con valor p de 0.0085

9.6.3 Reflexión sobre las suposiciones y sensitividad

Antes de pasar a la siguiente sección, puede ser benéfico el detenerse un momento y reflexionar sobre las suposiciones que se han formulado con respecto a las pruebas de hipótesis estadísticas sobre las medias. Se ha hecho énfasis con anterioridad, en que los procedimientos inferenciales estadísticos proporcionan un camino objetivo y veraz para formular inferencias con respecto a las características de la población con base en muestras aleatorias. Estos procesos por lo general tienen éxito sólo cuando las suposiciones que se han formulado para el desarrollo de las distribuciones de muestreo apropiadas se adhieren en forma razonable a la población. Los enfoques fortuitos y casuales para la aplicación de los métodos estadísticos, sin una comprensión de sus suposiciones y de las posibles consecuencias si éstas no se satisfacen, muchas veces lleva a una mala interpretación y a conclusiones erróneas.

Como ya se ha visto, la distribución t de Student juega un papel muy importante para formular inferencias con respecto a las medias, en forma especial en muestras de tamaño modesto. Pero la distribución t se basa en la suposición de que el muestreo se lleva a cabo sobre una distribución normal. Si el muestreo no se lleva a efecto sobre una distribución normal, el uso de la distribución t de Student es incorrecto debido a que, por ejemplo, las regiones críticas de tamaño α son probablemente más grandes que el valor que se especifica para α. Sin embargo, en forma afortunada, la distribución t es muy ***robusta,*** o insensible a la suposición de normalidad, y en forma especial cuando el tamaño de la muestra es mayor o igual a 15.

Cuando se emplea la distribución t de Student para comparar dos medias, es mucho más severo violar la suposición de varianzas iguales que la suposición de normalidad. Por una razón intuitiva del efecto aparente, supóngase que en realidad se están muestreando dos distribuciones normales, una con media $\mu = 100$ y desviación estándar $\sigma = 20$, y la otra con $\mu = 120$ y $\sigma = 30$. El intervalo cuatro sigma de la primera es de 60 a 140 mientras que para la segunda es de 60 a 180. Por lo tanto, puede observarse un valor menor o igual a 140 en cualquiera de las dos poblaciones. Sin embargo, estos valores no implicarán que exista una diferencia entre

las dos medias. Únicamente las observaciones de una segunda muestra que sean mayores de 140 empezarían a sugerir una diferencia media aparente, pero su número es probablemente demasiado pequeño para hacer la diferencia entre las medias discernibles. De esta forma, con base en la estadística T es probable que se llegue a la conclusión equivocada de que no existe diferencia entre las medias con una frecuencia inaceptable debida al desbalance en la variación inherente de las dos distribuciones.

Para cuantificar el efecto de varianzas desiguales se simularon 1 000 muestras aleatorias, cada una de tamaño 20 a partir de dos distribuciones normales mediante el empleo de paquete IMSL. Para la primera distribución se escogieron los valores de la media y de la desviación estándar iguales a 100 y 20, respectivamente. Para la segunda se emplearon los valores de 110, 120 y 130 para la media, y los valores de 25, 30 y 40 para la desviación estándar. De acuerdo con lo anterior se simularon 12 casos donde para cada par de muestras aleatorias se probó la hipótesis

$$H_0: \mu_1 - \mu_2 = 0$$

contra la alternativa

$$H_1: \mu_1 - \mu_2 < 0$$

mediante el uso de la estadística T de Student dada por (9.15). Para cada caso se determinó el número, de entre 1 000 ensayos, para el que la hipótesis nula no podía rechazarse con $\alpha = 0.05$. De esta forma es posible comparar el tamaño del error de tipo II para cada caso contra el valor correspondiente que puede obtenerse de las curvas CO en [1], cuando ambas desviaciones estándar tienen un valor igual a 20. Las probabilidades para el error de tipo II se dan en la tabla 9.10. Cuando se comparan los valores β para varianzas iguales, existe un incremento apreciable en el tamaño del error de tipo II conforme la diferencia entre las varianzas es más pronunciada. Por lo tanto, el efecto de violar la suposición de varianzas iguales cuando se comparan las medias puede ser sustancial.

Ahora se examinará el efecto en el tamaño del error de tipo I si se viola la suposición de varianzas iguales. Esto es, si se supone que H_0 es cierta, ¿qué efecto pueden tener las varianzas desiguales sobre α? Scheffé [4] determinó que si los tamaños de las muestras n_1 y n_2 son grandes pero iguales, la estadística T es considerablemente más robusta a la suposición de varianzas iguales cuando se comparan dos medias. La tabla 9.11 (véase [4] para los detalles) contiene el tamaño del error de tipo I con

TABLA 9.10 Probabilidades β simuladas para el efecto de varianzas desiguales cuando se comparan dos medias ($\mu_1 = 100$, $\sigma_1 = 20$)

	$\sigma_2 = 20$	$\sigma_2 = 25$	$\sigma_2 = 30$	$\sigma_2 = 40$
$\mu_2 = 110$	0.550	0.626	0.687	0.758
$\mu_2 = 120$	0.065	0.139	0.209	0.389
$\mu_2 = 130$	0.002	0.008	0.021	0.093

TABLA 9.11 Probabilidades α para el efecto de varianzas desiguales cuando se comparan dos medias

		σ_1^2/σ_2^2				
		1/5	1/2	1	2	5
	1	0.050	0.050	0.050	0.050	0.050
n_1/n_2	2	0.120	0.080	0.050	0.029	0.014
	5	0.220	0.120	0.050	0.014	0.002

base en un intervalo de confianza del 95% para $\mu_1 - \mu_2$ como una función del cociente de los dos tamaños muestrales y el cociente de las dos varianzas. Nótese que el tamaño del error de tipo I no cambia en el primer renglón con respecto a su valor preestablecido de 0.05, aun a pesar de que el cociente de las varianzas cambie.

A través de toda la discusión de la inferencia estadística se ha supuesto que se obtiene una muestra aleatoria y que por lo tanto las observaciones se encuentran independientemente distribuidas. Si estas suposiciones no se cumplen, es probable que cualquier inferencia estadística que se formule sea errónea sin importar el tamaño de la muestra. Aún así, la suposición que, en forma probable, es la que se viola, la mayoría de las veces es la de una muestra aleatoria.

Relacionado en forma cercana al concepto de aleatoriedad, es la selección de la muestra cuando las medias de los dos niveles (o más, como se estudiará mas adelante) se comparan entre sí. Con propósitos de ilustración, recuérdese el ejemplo 9.9. Dado que se seleccionaron 16 personas aleatoriamente para desempeñar la tarea dada bajo el nivel 1, se deduce que las personas que realizaron la tarea en el nivel 2 también fueron seleccionadas de manera aleatoria. Este procedimiento asegura una asignación imparcial de cuáles de las 32 personas se encontrarán sujetas a un determinado nivel de ruido. En inferencia estadística este proceso de selección imparcial recibe el nombre de *aleatorización*. El principio de aleatorización protege contra la introducción de sesgo sistemático en la asignación de personas u objetos a diferentes niveles y por ello consolida la credibilidad de la inminente comparación.

Se ha visto cómo las diferencias inherentes en la variabilidad pueden oscurecer la comparación entre dos medias. Muchas veces, durante el proceso de observar datos muestrales, factores externos no controlados pueden causar diferencias en la variabilidad. Sin embargo, mediante la adhesión al principio de aleatorización, estos factores externos probablemente tengan un efecto balanceado sobre las mediciones bajo los dos niveles de interés. Por ejemplo, en el problema del ruido, factores tales como el estado de ánimo del individuo en el momento de realizar la tarea no pueden ser controlados. El principio de aleatorización tiende a neutralizar tales efectos.

9.6.4 Prueba sobre las medias cuando las observaciones están pareadas

De la última sección recuérdese que cuando se comparan las medias de dos niveles, es deseable tener a las personas u objetos que producirán las observaciones dentro de

cada nivel, tan homogéneas como sea posible. Si existe un efecto debido a factores externos, éstos pueden neutralizarse mediante la aplicación del principio de aleatorización. También es posible controlar la variación no deseada controlando los factores extraños. Esto se logra tomando las observaciones en pares, donde se supone que las condiciones externas son las mismas para cada par pero pueden variar de par en par. En forma general, existe una relación natural entre las observaciones de un par. Esto es, para cada par se selecciona una persona u objeto al azar y se somete a ambos niveles de interés. A pesar de que se desea determinar si existe alguna diferencia entre las medias, no puede considerarse a los pares como dos muestras aleatorias independientes.

Como ilustración, se examinará el siguiente problema: un investigador médico se interesa en determinar si un fármaco experimental tiene el efecto colateral no deseable de elevar la presión sistólica sanguínea. Para conducir un estudio de amplia cobertura se seleccionan en forma aleatoria *n* personas de diferentes edades y condiciones de salud. En un ambiente controlado de laboratorio se toma la presión sanguínea de los *n* sujetos y se les administra el fármaco durante un lapso adecuado de tiempo después del cual se les vuelve a tomar la presión sanguínea.

Sean $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ los n pares, donde (X_i, Y_i) denota la presión sistólica sanguínea del i-ésimo sujeto antes y después de adiministrar el medicamento. Nótese que en este caso los factores externos son la condición del individuo en relación con su edad, su salud y otras pecualiaridades que pueden tener un efecto único sobre la presión sanguínea. Puesto que cada sujeto forma un par, el efecto de los factores externos sobre la presión sanguínea se encuentra entre los pares y cualquier diferencia sustancial de la presión dentro de cada par puede atribuirse al efecto de la droga. Así, al tomar la diferencia entre las dos observaciones de cada par es posible remover (bloquear) la variabilidad en la presión sanguínea a consecuencia de los factores externos. Esto hace posible una comparación válida de la presión sanguínea antes y después de administrar el medicametno. Por lo tanto, el interés se centra en la columna de diferencias de la tabla 9.12 generada al restar una observación de la otra para cada par.

Se supone que las diferencias $D_1, D_2, \ldots, D_n$ constituyen variables aleatorias independientes distribuidas normales tales que $E(D_i) = \mu_D$ y $Var(D_i) = \sigma_D^2$ para toda $i = 1, 2, \ldots, n$. Lo anterior es posible si se supone independencia entre los pa-

TABLA 9.12 Diferencias entre las observaciones en un experimento

Número de par (persona)	*Nivel 1* (PS *antes*)	*Nivel 2* (PS *después*)	*Diferencia* $Y - X$*
1	X_1	Y_1	$D_1 = Y_1 - X_1$
2	X_2	Y_2	$D_2 = Y_2 - X_2$
.	.	.	.
.	.	.	.
.	.	.	.
n	X_n	Y_n	$D_n = Y_n - X_n$

* Puede tomarse fácilmente la diferencia $X - Y$.

res (pero no necesariamente entre los valores de éstos) de manera tal que $E(X_i) = \mu_i$ y $E(Y_i) = \mu_i + \mu_D$ para $i = 1, 2 \ldots n$. De esta forma para el i-ésimo par, los valores esperados difieren por una constante, la cual es el valor esperado de D_i para $i = 1, 2 \ldots n$. Además, $Var(X_i) = \sigma_X^2$ y $Var(Y_i) = \sigma_Y^2$ son desconocidas y no necesariamente iguales, pero se supone que son constantes para toda $i = 1, 2, \ldots, n$.

En el contexto del problema de la presión sanguínea, lo que se está diciendo es lo siguiente: la constante μ_D es la diferencia media en la presión sanguínea como consecuencia del medicamento. Aun a pesar de que las presiones sanguíneas promedio varían de persona a persona por las diferencias en las condiciones de salud, se piensa que μ_D es probablemente igual para todas las personas. Nótese que si μ_D fuese cero, esto podría sugerir que el medicamento no tiene ningún efecto sobre la presión sanguínea. Por otro lado, si μ_D es mayor que cero, esto podría indicar un incremento de la presión sanguínea promedio a consecuencia del medicamento. La varianza σ_D^2 de las diferencias en la presión sanguínea no es conocida y depende de las varianzas antes y después de administrarse el medicamento. A pesar de que las varianzas σ_X^2 y σ_Y^2 pueden ser diferentes, se supone que son constantes de persona a persona.

La discusión anterior demuestra que se pueden formular inferencias sobre las medias de dos niveles cuando las observaciones están pareadas al considerar la columna de diferencias como una sola variable aleatoria y al aplicar los métodos de la sección 9.6.1. Bajo la hipótesis nula

$$H_0: \mu_D = \delta_0,$$

la estadística

$$T = \frac{\overline{D} - \delta_0}{S_D/\sqrt{n}} \tag{9.16}$$

tiene una distribución t de Student con $n - 1$ grados de libertad, en donde

$$\overline{D} = \sum_{i=1}^{n} D_i/n$$

y

$$S_D^2 = \sum_{i=1}^{n} (D_i - \overline{D})^2/(n - 1).$$

Las regiones críticas de tamaño α para las hipótesis alternativas uni y bilaterales se encuentran resumidas en la tabla 9.13.

Ejemplo 9.10 En el problema anterior de la presión sanguínea, sea $\alpha = 0.01$ y pruébese la hipótesis nula

$$H_0: \mu_D = 0$$

contra la alternativa

$$H_1: \mu_D > 0,$$

TABLA 9.13 Criterios de rechazo para la prueba de hipótesis con respecto a las medias cuando las observaciones están pareadas

Hipótesis nula	*Valor de la estadística de prueba bajo* H_0
$H_0: \mu_D = \delta_0$	$t = \dfrac{\bar{d} - \delta_0}{s_d/\sqrt{n}}$
Hipótesis alternativa	***Criterios de rechazo***
$H_1: \mu_D \neq \delta_0$	Rechazar H_0 cuando $t \leq t_{\alpha/2,\, n-1}$ o cuando $t \geq t_{1-\alpha/2,\, n-1}$
$H_1: \mu_D > \delta_0$	Rechazar H_0 cuando $t \geq t_{1-\alpha,\, n-1}$
$H_1: \mu_D < \delta_0$	Rechazar H_0 cuando $t \leq t_{\alpha,\, n-1}$

con base en los datos muestrales de la tabla 9.14.

En la columna de diferencias se tiene que $\bar{d} = 3.75$ y $s_D = 3.7929$. De esta forma el valor de la estadística de prueba es

$$t = \frac{3.75 - 0}{3.7929/\sqrt{12}} = 3.425.$$

Dado que el valor crítico es $t_{0.99,\, 11} = 2.718$, se rechaza la hipótesis nula de no efecto del medicamento. Por lo tanto, con base en los resultados de este estudio, un incremento en el valor promedio de la presión sanguínea es estadísticamente discernible con un valor p de 0.0036.

Es importante notar que en el ejemplo anterior no existe ninguna oportunidad de aplicar el principio de aleatorización para remover los posibles sesgos sistemáticos.

TABLA 9.14 Datos muestrales para el ejemplo 9.10

Sujeto	PS *antes*	PS *después*	*Diferencias (después − antes)*
1	128	134	6
2	176	174	−2
3	110	118	8
4	149	152	3
5	183	187	4
6	136	136	0
7	118	125	7
8	158	168	10
9	150	152	2
10	130	128	−2
11	126	130	4
12	162	167	5

Lo anterior es típico de las situaciones antes-después en las que las observaciones se aparean con el propósito de remover efectos externos. Sin embargo, es posible que intervengan otros factores externos entre las mediciones y que éstos causen diferencias sustanciales en las observaciones de algunos pares; esta influencia será acreditada de manera equivocada a los efectos que se están verificando. En el problema de la presión sanguínea algunos de los sujetos pueden sufrir cambios en su salud que sean independientes del medicamento que se les administra, y estos cambios pueden a su vez causar un aumento (o disminución) de la presión sanguínea. El siguiente ejemplo proporciona un experimento mejor para comparar dos medias para observaciones pareadas.

Ejemplo 9.11 La investigación ha desarrollado variedades superiores de maíz que proporcionarán cantidades más grandes de éste por unidad de tierra. Un investigador ha desarrollado una nueva variedad híbrida de este grano y piensa que es superior a la mejor variedad disponible. También cree que esta nueva variedad rebasará con mucho la producción estándar en varias localidades geográficas. Para verificar lo anterior, el investigador diseña el siguiente experimento: se seleccionan 10 parcelas de igual tamaño cada una en distinta localidad geográfica. Cada parcela se divide en dos secciones iguales, de manera tal que puedan cultivarse las dos variedades en cada localidad. Para remover los posibles sesgos sistemáticos, se aplica el principio de aleatorización a todas las parcelas para decidir qué sección es la que se cultiva y con qué tipo de variedad. Lo anterior se logra lanzando una moneda para decidir la variedad. Se controlan tantos factores como es posible; por ejemplo, la temporada de siembra, el tipo de fertilizante y el intervalo de aplicación. En el momento de recoger la cosecha, se anotan las toneladas por unidad de área. Supóngase que los datos que se muestran en la tabla 9.15 son los que se observaron. Con base en estos datos, obténgase un intervalo de confianza del 95% para la diferencia media en la producción entre las variedades X y Y.

Antes de proceder con el análisis, debe notarse que se están bloqueando los factores externos como resultado del apareamiento en la localidad geográfica. En situaciones de este tipo, existe muy poca duda con respecto a que las condiciones de la tierra y otros efectos probablemente no sean los mismos en las diferentes localidades. De esta forma existe una gran oportunidad para observar un efecto sustancial sobre la producción a consecuencia de la localidad. También, nótese que esta oportunidad se presenta al aleatorizar la asignación de variedades a las parcelas para remover cualquier sesgo sistemático.

TABLA 9.15 Datos muestrales para el ejemplo 9.11

Tipo	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9	L_{10}
Variedad X (estándar)	23	35	29	42	33	19	37	24	35	26
Variedad Y (nueva)	26	39	35	40	38	24	36	27	41	27

Para obtener el intervalo de confianza deseado, las diferencias entre las producciones de X y Y en las 10 localidades son -3, -4, -6, 2, -5, -5, 1, -3, -6, y -1. Con base en éstas, $\bar{d} = -3$ y $s_D = 2.8284$. Asumiendo que estas diferencias son los valores de dos variables aleatorias independientes y normalmente distribuidas, un intervalo de confianza del 95% para μ_D es

$$\bar{d} \pm t_{0.975,\,9} \frac{s_D}{\sqrt{10}},$$

o

$$-3 \pm (2.262)(2.8284/\sqrt{10}),$$

el que se reduce al intervalo $(-5.0232, -0.9768)$. Dado que el valor cero no se incluye en este intervalo, se rechaza la correspondiente hipótesis nula de que la diferencia es cero a un nivel de $\alpha = 0.05$

Resulta apropiado colocar el problema de comparar las medias de dos niveles en una mejor perspectiva para justificar la planeación de un experimento con base en muestras independientes o con base en muestras pareadas. Sean X y Y los dos niveles de interés, asumiendo un tamaño n igual para las dos muestras independientes y n pares de observaciones. Dado que lo que se desea en cualquiera de los casos es una inferencia con respecto a la diferencia entre las medias, la estadística para ambos casos es $\bar{X} - \bar{Y}$. De esta manera, bajo la suposición de que se muestrean distribuciones normales un intervalo de confianza del $100(1 - \alpha)\%$ para la diferencia media en cualquiera de los casos es de la forma general

$$(\bar{X} - \bar{Y}) \pm t_{1-\alpha/2,\,m}\, d.e.(\bar{X} - \bar{Y}), \tag{9.17}$$

donde m es el número de grados de libertad. En la expresión (9.17) existen dos términos que difieren en ambos casos. Uno es el valor cuantil $t_{1-\alpha/2,m}$; y el otro es la desviación estándar de la estadística $\bar{X} - \bar{Y}$. Cuando las observaciones son pareadas, el valor cuantil es una función de $m = n - 1$ grados de libertad, mientras que para muestras independientes se basa en $m = 2(n - 1)$ grados de libertad. Para un α dado, el valor cuantil aumenta conforme el número de grados de libertad disminuye. Entonces, un intervalo de confianza para muestras pareadas es más amplio debido a la pérdida de grados de libertad.

A la luz de la información anterior, la desviación estándar de $\bar{X} - \bar{Y}$ se convierte en un cambio a mantener en mente cuando se escoge entre muestras independientes o muestras pareadas. Si se permite a un factor extraño, el cual influye en forma potencial que varíe, cuando se toman las muestras independientes, la consecuencia probable es una variabilidad importante entre las observaciones, dando como consecuencia un valor grande $d.e.(\bar{X} - \bar{Y})$. Al parear las observaciones, es posible neutralizar la influencia del factor extraño y mantener su efecto igual dentro de cada par. Entonces, las observaciones dentro de cada par estarán probablemente correlacionadas. Esto es, para un par dado, es probable que un valor grande de X dé como resultado un valor grande de Y o viceversa, lo cual da como resultado una covarianza positiva entre X y Y. Se sigue entonces que, dado $Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$, la varianza de $X - Y$ (así como también la de $\bar{X} - \bar{Y}$)

será más pequeña para muestras pareadas que para muestras independientes. Por lo tanto, en un experimento bien planeado para observaciones pareadas, la reducción en el valor de la desviación estándar de $\bar{X} - \bar{Y}$, por lo general compensará el aumento en el valor crítico debido a la reducción en el número de grados de libertad.

Como ilustración, en el ejemplo 9.11 se calculó el estimador $s_D = 2.8284$. Si los datos se consideran como muestras independientes de dos distribuciones normales con varianzas iguales, un estimado de la varianza común es

$$s_p^2 = \frac{9(52.6778) + 9(43.1222)}{18} = 47.9,$$

o $s_p = 6.921$ el valor $s_p = 6.921$ es más del doble del valor $s_D = 2.8284$. Al construir un intervalo de confianza del 95% para muestras independientes, se obtiene

$$-3 \pm (2.101)(6.921)\sqrt{\frac{1}{10} + \frac{1}{10}},$$

o

$$(-9.5029, 3.5029).$$

El obvio que no puede rechazarse la hipótesis nula de no diferencia entre las medias, si los datos fuesen considerados como muestras independientes.

9.7 Pruebas de hipótesis con respecto a las varianzas cuando se muestrean distribuciones normales

Se argumentó con anterioridad, que una inferencia con respecto a una varianza es tan importante como una con respecto a la media. En medios industriales, por ejemplo, la variabilidad de un producto puede ser una medida más importante que el promedio del producto. Por esta razón, así como también por la necesidad de comprobar la hipótesis de varianzas iguales, se presentarán criterios para probar hipótesis con respecto a las varianzas con base en una sola muestra aleatoria o con base en dos muestras aleatorias independientes provenientes de distribuciones normales. Como era de esperarse, los criterios para probar hipótesis con respecto a las varianzas se basan en los correspondientes métodos para construir intervalos de confianza, tal como se descutió en las secciones 8.4.4. y 8.4.5. Nuevamente es imperativo hacer énfasis en que estos procedimientos son, en forma especial, sensibles a la suposición de normalidad.

9.7.1 Puebas para una muestra

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución normal con media μ desconocida y varianza σ^2 desconocida. Considérese nula la prueba de la siguiente hipótesis

$$H_0\colon \sigma^2 = \sigma_0^2$$

contra una de las siguientes alternativas

$$H_1: \sigma^2 \neq \sigma_0^2, \qquad H_1: \sigma^2 > \sigma_0^2, \qquad H_1: \sigma^2 < \sigma_0^2,$$

donde σ_0^2 es el valor propuesto para σ^2. La estadística de interés es la varianza muestral S^2. La hipótesis nula será rechazada si la realización de S^2 calculada a partir de la muestra, es, en forma suficiente, diferente, mayor que o menor que σ_0^2, dependiendo de la hipótesis alternativa. Pero bajo H_0, la cantidad $(n - 1)s^2/\sigma_0^2$ es un valor de una variable aleatoria chi-cuadrada con $n - 1$ grados de libertad. Entonces, por ejemplo, si la hipótesis alternativa es H_1: $\sigma^2 > \sigma_0^2$, se rechazará a H_0 si el valor de $(n - 1)s^2/\sigma_0^2$ se encuentra dentro de la región crítica de tamaño α en el lado derecho de la distribución chi-cuadrada con $n - 1$ grados de libertad. En la tabla 9.16 se proporciona la información más relevante al respecto.

Como se notó con anterioridad, la violación de la suposición de que el muestreo se lleva a cabo sobre una distribución normal tiene un efecto sustancial cuando se emplea la estadística chi-cuadrada para inferencias con respecto a las varianzas. Para ilustrar este efecto, se simuló un experimento similar al descrito en la sección 8.4.3. Para un tamaño de la muestra $n = 30$, se generaron 1 000 muestras aleatorias para cada una de las siguientes distribuciones: uniforme, exponencial y gama. Los valores de los parámetros de cada distribución se seleccionaron en cada caso para proporcionar una varianza de 100. Para cada muestra aleatoria se probó la hipótesis nula

$$H_0: \sigma^2 = 100$$

contra la alternativa

$$H_1: \sigma^2 > 100,$$

mediante el empleo de la estadística chi-cuadrada con $\alpha = 0.05$. Para cada distribución se contó el número de veces para las que se rechazaba la hipótesis nula. Los resultados se encuentran en la tabla 9.17.

Dado que $\alpha = 0.05$ representa la probabilidad de rechazar una hipótesis cierta (tal cual es el caso aquí), se espera que 50 de las 1 000 muestras proporcionen esta de-

TABLA 9.16 Criterios de rechazo para la prueba de hipótesis con respecto a la varianza de una distribución normal con media desconocida

Hipótesis nula	*Valor de la estadística de prueba bajo H_0*
H_0: $\sigma^2 = \sigma_0^2$	$\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$
Hipótesis alternativa	*Criterios de rechazo*
H_1: $\sigma^2 \neq \sigma_0^2$	Rechazar H_0 cuando $\chi^2 \geq \chi^2_{1-\alpha/2,\, n-1}$, o cuando $\chi^2 \leq \chi^2_{\alpha/2,\, n-1}$
H_1: $\sigma^2 > \sigma_0^2$	Rechazar H_0 cuando $\chi^2 \geq \chi^2_{1-\alpha,\, n-1}$
H_1: $\sigma^2 < \sigma_0^2$	Rechazar H_0 cuando $\chi^2 \leq \chi^2_{\alpha,\, n-1}$

TABLA 9.17 Número de rechazos de la hipótesis nula de entre 1 000 muestras para tres distribuciones de igual varianza

Tipo de distribución y valores de los parámetros		
Uniforme $(0, \sqrt{1200})$	*Gama* *Forma* = 2; *Escala* = $\sqrt{50}$	*Exponencial* *Media* = 10
8	107	156

cisión cuando se muestree una distribución normal. Sin embargo, con base en los resultados existe una discrepancia suficiente para creer que la estadística chi-cuadrada es sensible a la suposición de que el muestreo se lleva a cabo sobre una distribución normal. No está por demás notar que los resultados del estudio de simulación son de alguna manera predecibles, especialmente si se comparan los factores de forma de las distribuciones seleccionadas con los de la distribución normal. La distribución uniforme es simétrica, al igual que la normal, pero se encuentra definida en el intervalo $(0, \sqrt{1200})$. Como consecuencia, la verosimilitud disminuye porque algunas muestras pueden contener valores extremos que pueden aumentar el valor de la varianza muestral. Así, el número de rechazos es menor que el que se espera. La distribución exponencial es la que tiene una mayor asimetría de entre las tres distribuciones seleccionadas y el mayor valor de curtosis. Por lo tanto, no es sorprendente que el número de rechazos sea mucho más grande que el correspondiente a una distribución normal. La distribución gama, con parámetros de forma y escala iguales a 2 y $\sqrt{50}$, respectivamente, se encuentra entre las anteriores ya que su coeficiente de asimetría es $\sqrt{2}$ y su curtosis relativa es 6.

9.7.2 Pruebas para dos muestras

Sean $X_1, X_2, \ldots, X_{n_X}$ y $Y_1, Y_2, \ldots, Y_{n_Y}$ dos muestras aleatorias de dos distribuciones normales independientes con medias desconocidas μ_X y μ_Y y varianzas desconocidas σ_X^2 y σ_Y^2. Considérese la prueba de la siguiente hipótesis nula

$$H_0\colon \sigma_X^2 = \sigma_Y^2$$

contra una de las siguientes alternativas:

$$H_1\colon \sigma_X^2 \neq \sigma_Y^2, \qquad H_1\colon \sigma_X^2 > \sigma_Y^2, \qquad H_1\colon \sigma_X^2 < \sigma_Y^2.$$

Las estadísticas de interés son las varianzas muestrales S_X^2 y S_Y^2. Por ejemplo, con respecto a la hipótesis alternativa bilateral, puede rechazarse la hipótesis nula si el estimador s_X^2 es lo suficientemente diferente del estimador s_Y^2. De la sección 7.8, recuérdese que por virtud de la independencia, las cantidades $(n_X - 1)S_X^2/\sigma_X^2$ y $(n_Y - 1)S_Y^2/\sigma_Y^2$ son dos variables aleatorias independientes chi-cuadrada con $n_X - 1$ y $n_Y - 1$ grados de libertad, respectivamente. Entonces se sigue la estadística

$$F = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2}$$

tiene una distribución F con $n_X - 1$ y $n_Y - 1$ grados de libertad. Pero bajo la hipótesis nula, $\sigma_X^2 = \sigma_Y^2$; de esta forma la estadística se reduce a

$$F = S_X^2/S_Y^2.$$

Para una hipótesis alternativa bilateral y un tamaño α del error de tipo I, se rechazará la hipótesis nula cuando $f = s_X^2/s_Y^2 \geq f_{1-\alpha/2,\, n_X-1,\, n_Y-1}$ o cuando $f \leq 1/f_{1-\alpha/2,\, n_Y-1,\, n_X-1}$. En la tabla 9.18 se proporciona un resumen completo de los criterios de rechazo.

Como ilustración, recuérdese que en el ejemplo 9.9, se asumió que las varianzas eran iguales al comparar las medias para los dos niveles de ruido. Para verificar la validez de esta suposición a un nivel de $\alpha = 0.1$, supóngase que se prueba la hipótesis

$$H_0\colon \sigma_1^2 = \sigma_2^2$$

contra la alternativa

$$H_1\colon \sigma_1^2 \neq \sigma_2^2.$$

Se observa que los valores críticos, izquierdo y derecho, son $f_{0.95,\,15,\,15} = 2.40$ y $1/f_{0.95,\,15,\,15} = 1/2.40 = 0.42$, respectivamente. Con base en los datos de la muestra $s_1^2 = 5.1833$ y $s_2^2 = 6.0$. De esta forma el valor de la estadística de prueba es

$$f = 5.1833/6 = 0.8639.$$

Dado que $f = 0.8639$ no es ni mayor ni igual a 2.4, ni menor o igual a 0.42, no es posible rechazar la hipótesis nula. De acuerdo con lo anterior, los resultados muestrales no proporcionan una razón válida para sospechar que está siendo violada la suposición de varianzas iguales.

TABLA 9.18 Criterios de rechazo para la prueba de hipótesis con respecto a las varianzas de dos distribuciones normales independientes

Hipótesis nula	*Valor de la estadística de prueba bajo H_0*
$H_0\colon \sigma_X^2 = \sigma_Y^2$	$f = s_X^2/s_Y^2$
Hipótesis alternativa	*Criterios de rechazo*
$H_1\colon \sigma_X^2 \neq \sigma_Y^2$	Rechazar H_0 cuando $f \geq f_{1-\alpha/2,\, n_X-1,\, n_Y-1}$, o cuando $f \leq 1/f_{1-\alpha/2,\, n_Y-1,\, n_X-1}$
$H_1\colon \sigma_X^2 > \sigma_Y^2$	Rechazar H_0 cuando $f \geq f_{1-\alpha,\, n_X-1,\, n_Y-1}$
$H_1\colon \sigma_X^2 < \sigma_Y^2$	Rechazar H_0 cuando $f \leq 1/f_{1-\alpha,\, n_Y-1,\, n_X-1}$

9.8 Inferencias con respecto a las proporciones de dos distribuciones binomiales independientes

En la sección 8.4.6 se desarrollaron los criterios para la construcción de intervalos de confianza para el parámetro de proporción p, cuando se muestrea una distribución binomial. En muchas ocasiones, el interés recae en comparar la proporción de un grupo distinto con la de otro, en relación con alguna característica en común. Por ejemplo, puede tenerse interés en comparar la proporción de unidades defectuosas para un producto dado, que se fabricó por dos compañías que compiten entre sí. O puede existir algún interés en comparar las proporciones de estudiantes de preparatoria en dos localidades geográficas diferentes que tienen un número de respuestas correctas para la prueba SAT por encima de cierto nivel. De esta forma, es necesario entender las ideas presentadas en la sección 8.4.6 para comparar los parámetros de proporción cuando se muestrean dos distribuciones binomiales independientes.

Como ilustración, en un estudio reciente se compararon las proporciones de personas zurdas y derechas que fuman. La población general se dividió en dos grupos, zurdos y derechos, y cada grupo fue subdividido en fumadores y no fumadores. Sea p_1 la proporción de personas zurdas que fuman y p_2 la proporción de personas derechas que fuman. El interés recae en hacer una comparación entre p_1 y p_2.

Supóngase que los zurdos y los derechos constituyen dos distribuciones binomiales independientes tales que la proporción de fumadores en los dos grupos es p_1 y p_2, respectivamente. Con base en muestras aleatorias de tamaño n_1 y n_2, sean X y Y el número observado de personas zurdas y derechas que fuman, respectivamente. Las proporciones muestrales

$$\hat{P}_1 = X/n_1,$$

$$\hat{P}_2 = Y/n_2$$

son los estimadores de máxima verosimilitud de p_1 y p_2, respectivamente. Dado que por hipótesis X y Y son variables aleatorias binomiales, las varianzas de los estimadores están dadas por

$$Var(\hat{P}_1) = Var(X/n_1) = p_1(1 - p_1)/n_1,$$

$$Var(\hat{P}_2) = Var(Y/n_2) = p_2(1 - p_2)/n_2.$$

Supóngase que se desea construir un intervalo de confianza muestral grande para la diferencia entre p_1 y p_2. La estadística de interés es la diferencia entre las dos proporciones muestrales. Ya que

$$E(\hat{P}_1) = p_1, \qquad E(\hat{P}_2) = p_2,$$

entonces, con base en el teorema 6.1 y su corolario dado por la expresión (7.2)

$$E(\hat{P}_1 - \hat{P}_2) = p_1 - p_2, \tag{9.18}$$

y

$$Var(\hat{P}_1 - \hat{P}_2) = Var(\hat{P}_1) + Var(\hat{P}_2)$$
$$= \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}. \tag{9.19}$$

Con base en una discusión anterior (véase el capítulo 5) puede demostrarse que en valores grandes de n_1 y n_2, la distribución de la estadística $\hat{P}_1 - \hat{P}_2$ es, en forma aproximada, normal con media y varianza dadas por (9.18) y (9.19), respectivamente. En otras palabras, la distribución de

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)}{\sqrt{\dfrac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \dfrac{\hat{P}_2(1-\hat{P}_2)}{n_2}}} \tag{9.20}$$

es aproximadamente $N(0,1)$ n_1 y n_2. Nótese que el denominador en la expresión (9.20) proporciona un estimador de la desviación estándar de la estadística $\hat{P}_1 - \hat{P}_2$, ya que se han reemplazado las proporciones muestrales p_1 y p_2. Por lo tanto, se sigue que para n_1 y n_2 grandes, la probabilidad del intervalo aleatorio

$$[(\hat{P}_1 - \hat{P}_2) - z_{1-\alpha/2}\, d.e.(\hat{P}_1 - \hat{P}_2), \quad (\hat{P}_1 - \hat{P}_2) + z_{1-\alpha/2}\, s.d.(\hat{P}_1 - \hat{P}_2)]$$

es aproximadamente $1 - \alpha$, y un intervalo de confianza aproximado del $100(1 - \alpha)\%$ para $p_1 - p_2$ es:

$$(\hat{p}_1 - \hat{p}_2) \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}, \tag{9.21}$$

en donde $\hat{p}_1 = x/n_1$ y $\hat{p}_2 = y/n_2$ son los estimados de máxima verosimilitud de p_1 y p_2 respectivamente.

Ejemplo 9.12 En un estudio de los hábitos de fumador para personas zurdas y derechas, una muestra aleatoria de 400 zurdos reveló que 190 de éstos fuman, y en una muestra aleatoria de 800 derechos, 300 de éstos fuman. Con base en esta evidencia, construir un intervalo de confianza del 98% para la diferencia real entre las proporciones p_1 y p_2.

Los estimados de las proporciones son

$$\hat{p}_1 = 190/400 = 0.475, \qquad \hat{p}_2 = 300/800 = 0.375.$$

Dado que los tamaños de las muestras son grandes, la aproximación normal es adecuada para este caso. Para un intervalo de confianza del 98% $z_{0.99} = 2.33$ y el intervalo de confianza es

$$(0.475 - 0.375) \pm 2.33\sqrt{\frac{(0.475)(1-0.475)}{400} + \frac{(0.375)(1-0.375)}{800}},$$

el cual simplifica al intervalo (0.0295, 0.1705). Dado que este intervalo de confianza no incluye al origen y, de hecho, se encuentra a la derecha de éste, puede concluirse con un 98% de confiabilidad, que el porcentaje de zurdos que fuman es mayor que el correspondiente para las personas derechas.

Supóngase que el interés recae en probar la hipótesis nula

$$H_0: p_1 - p_2 = 0$$

contra una de las siguientes alternativas:

$$H_1: p_1 - p_2 \neq 0, \qquad H_1: p_1 - p_2 > 0, \qquad H_1: p_1 - p_2 < 0.$$

Dadas muestras aleatorias de tamaños n_1 y n_2, considérese la estadística $\hat{P}_1 - \hat{P}_2$. La intuición sugiere que debe rechazarse la hipótesis nula si un valor de la estadística es, en forma suficiente, diferente, mayor que, o menor que cero, dependiendo de la hipótesis alternativa. En forma equivalente, la decisión puede basarse en una prueba estadística similar a la dada por (9.20), la cual es aproximadamente $N(0, 1)$ para valores grandes de n_1 y n_2.

Dado que bajo H_0 se supone que las dos proporciones son iguales, sea $p = p_1 = p_2$ la proporción común. Entonces, si la hipótesis nula es cierta, la estadística $\hat{P}_1 - \hat{P}_2$ tiene una distribución, en forma aproximada, normal con media

$$E(\hat{P}_1 - \hat{P}_2) = 0$$

y desviación estándar

$$d.e.(\hat{P}_1 - \hat{P}_2) = \sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}}$$
$$= \left(\sqrt{p(1-p)}\right)\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right).$$

Ya que el valor de p no se conoce, se combina la información de las dos muestras para obtener el estimador combinado

$$\hat{P} = \frac{X + Y}{n_1 + n_2},$$

donde X y Y son las variables aleatorias que se observaron y que poseen la característica de interés. Entonces un estimado de la desviación estándar de $\hat{P}_1 - \hat{P}_2$ es

$$d.e.(\hat{P}_1 - \hat{P}_2) = \left(\sqrt{\hat{p}(1-\hat{p})}\right)\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right),$$

en donde $\hat{p} = (x + y)/(n_1 + n_2)$ es el estimador combinado de p. Bajo H_0 la estadística

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\left(\sqrt{\hat{P}(1-\hat{P})}\right)\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)} \tag{9.22}$$

es aproximadamente $N(0, 1)$ para valores grandes de n_1 y n_2. Dependiendo de la hipótesis alternativa, el lector no debe tener dificultad para decidir cuándo rechazar H_0 con base en (9.22) dado un tamaño del error de tipo I.

Referencias

1. A. H. Bowker and G. J. Lieberman, *Engineering statistics*, Prentice-Hall, Englewood Cliffs, N.J., 1959.
2. P. G. Hoel, *Introduction to mathematical statistics*, 4th ed., Wiley, New York, 1971.
3. B. W. Lindgren, *Statistical theory*, 3rd ed., Macmillan, New York, 1976.
4. H. Scheffé, *The analysis of variance*, Wiley, New York, 1959.

Ejercicios

9.1. Suponga que usted desea probar la hipótesis

$$H_0: \theta = 5$$

contra la alternativa

$$H_1: \theta = 8$$

por medio de un solo valor que se observa en una variable aleatoria con densidad de probabilidad $f(x; \theta) = (1/\theta)\exp(-x/\theta)$, $x > 0$. Si el tamaño máximo del error de tipo I que puede tolerarse es de 0.15, ¿cuál de las siguientes pruebas es la mejor para escoger entre las dos hipótesis?

a) Rechazar H_0 si $X \geqslant 9$
b) Rechazar H_0 si $X \geqslant 10$
c) Rechazar H_0 si $X \geqslant 11$

9.2. Suponga que usted observa un solo valor de una variable aleatoria cuya función de densidad está dada por $f(x; \theta) = 1/\theta$, $0 < x < \theta$, y desea probar la hipótesis

$$H_0: \theta = 20$$

contra la alternativa

$$H_1: \theta = 15.$$

¿cuál de las dos pruebas *a*) rechazar H_0 si $X \leqslant 8$, o *b*) rechazar H_0 si $X > 8$ es la mejor para decidir entre las dos hipótesis?

9.3. Se sabe que la proporción de artículos defectuosos en un proceso de manufactura es de 0.15. El proceso se vigila en forma periódica tomando muestras aleatorias de tamaño 20 e inspeccionando las unidades. Si se encuentran dos o más unidades defectuosas en la muestra, el proceso se detiene y se considera como "fuera de control".

a) Enunciar las hipótesis nula y alternativa apropiadas.
b) Obtener la probabilidad del error de tipo I.
c) Obtener y graficar la función de potencia para los siguientes valores alternativos de la proporción de artículos defectuosos: 0.06, 0.08, 0.1, 0.15, 0.2, y 0.25.

d) Compárense sus respuestas con las partes *b* y *c* para el caso en el que se juzga al proceso como fuera de control cuando se encuentran tres o más defectuosas.

9.4. La cantidad promedio que se coloca en un recipiente en un proceso de llenado se supone que es de 20 onzas. En forma periódica, se escogen al azar 25 recipientes y el contenido de cada uno de éstos se pesa. Se juzga al proceso como fuera de control cuando la media muestral $\bar{X}$ es menor o igual a 19.8 o mayor o igual a 20.2 onzas. Se supone que la cantidad que se vacía en cada recipiente se encuentra aproximada, en forma adecuada, por una distribución normal con una desviación estándar de 0.5 onzas.

a) Enúnciense las hipótesis nula y alternativa que son propias para esta situación.
b) Obtener la probabilidad del error de tipo I.
c) Obtener y graficar la función de potencia para los siguientes valores medios de llenado: 19.5, 19.6, 19.7, 19.8, 19.9, 20.0, 20.1, 20.2, 20.3, 20.4, y 20.5.
d) Como una prueba alternativa, considérese el rechazo de H_0 cuando $\bar{X} \leqslant 19.75$ o cuando $\bar{X} \geqslant 20.25$. Si el tamaño máximo del error de tipo I es de 0.05, ¿cuál de las dos pruebas es la mejor?

9.5. Con referencia al ejercicio 9.4, supóngase que el tamaño de la muestra se aumenta a 36 recipientes. Dados los mismos tamaños del error de tipo I para las pruebas propuestas, obtener los nuevos valores críticos y comparar las funciones potencia de las dos pruebas.

9.6. Los siguientes datos son los tiempos de sistema observados (tiempo de espera más tiempo de servicio) para 10 clientes en una tienda: 8.7, 2.4, 18.2, 10.5, 9.7, 4.8, 11.2, 29.3, 10.8, 15.6. Supóngase que el tiempo del sistema es una variable aleatoria con una distribución gama, con parámetro de forma igual a 2 y parámetro de escala θ desconocido. (Sugerencia: véase la expresión (5.51) y el teorema 7.1.)

a) Pruébese la hipótesis nula

$$H_0: \theta = 5$$

contra la alternativa

$$H_1: \theta > 5,$$

con un tamaño máximo del error de tipo I igual a 0.05.
b) Si el valor real de θ fuese 7, ¿cuál sería la probabilidad del error de tipo II?

9.7. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño n de una distribución normal con media μ desconocida y varianza σ^2 conocida. Obtener la mejor región crítica de tamaño α para probar

$$H_0: \mu = \mu_0$$

contra

$$H_1: \mu = \mu_1,$$

en donde $\mu_1 < \mu_0$.

9.8. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño n de una distribución de Poisson con parámetro λ desconocido. Obtener la mejor región crítica de tamaño α para probar

$$H_0: \lambda = \lambda_0$$

contra

$$H_1: \lambda = \lambda_1,$$

en donde $\lambda_1 < \lambda_0$.

9.9. El número de accidentes en un crucero muy transitado sigue el modelo exacto de una distribución de Poisson con una media de 2.5 accidentes por semana. Un ingeniero de tráfico decide reducir la velocidad límite de las dos avenidas que se intersectan en el crucero. La decisión con respecto a si la reducción en el límite de velocidad disminuye el número de accidentes promedio por semana, se tomará con base en el número total de accidentes que se observan durante un periodo de cuatro semanas a partir de la reducción en el límite de velocidad.

a) Enunciar las hipótesis nula y alternativa apropiadas para esta situación.
b) Para un tamaño máximo del error de tipo I igual a 0.1, obtener el valor crítico de la estadística de prueba para el rechazo de la hipótesis nula. (Sugerencia: véanse el ejemplo 9.4 y el ejercicio 7.6.)
c) Si el número de accidentes promedio disminuyó a 2, obtener la probabilidad del error de tipo II.

9.10. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño n de una distribución exponencial con parámetro de escala θ desconocido. Obtener la mejor región crítica de tamaño α para probar

$$H_0: \theta = \theta_0$$

contra

$$H_1: \theta = \theta_1,$$

donde $\theta_1 > \theta_0$.

9.11. Se seleccionaron al azar cuatro unidades de videojuegos y se probaron hasta que ocurre la falla de éstos. El tiempo que observaron los que tuvieron las fallas son 148.2, 120.6 165.5 y 145.7 horas. Supóngase que el lapso de tiempo que transcurre hasta que se presenta la falla es una variable aleatoria exponencial, empléese el ejemplo 7.4 para probar la hipótesis nula de que el tiempo medio para que una falla ocurra es de 140 contra la alternativa de que éste es mayor de 140 horas con una probabilidad del error de tipo I igual a 0.01. (Sugerencia: Empléese una técnica iterativa en conjunción con la expresión (5.56).)

9.12. Un contratista ordena un gran número de vigas de acero con longitud promedio de 5 metros. Se sabe que la longitud de una viga se encuentra normalmente distribuida con una desviación estándar de 0.02 metros. Después de recibir el embarque, el contratista selecciona 16 vigas al azar y mide sus longitudes. Si la media muestral tiene un valor más pequeño que el esperado, se tomará la decisión de enviar el embarque al fabricante.

a) Si la probabilidad de rechazar un embarque bueno es de 0.04, ¿cuál debe ser el valor de la media muestral para que el embarque sea regresado al fabricante?
b) Si la longitud promedio real es de 4.98 metros, ¿cuál es la potencia de la prueba en el inciso *a*?

9.13. En el ejercicio 9.12, ¿cuál es el tamaño necesario de la muestra para que la probabilidad de detectar una disminución de 0.015 metros en la longitud media sea de 0.99?

9.14. El propietario de una automóvil compacto sospecha que la distancia promedio por galón que ofrece su carro es menor que la especificada por la EPA, la cual es de 30 millas por galón. El propietario observa la distancia recorrida por galón en nueve ocasiones y obtiene los siguientes datos: 28.3, 31.2, 29.4, 27.2, 30.8, 28.7, 29.2, 26.5, 28.1. Después de una investigación el propietario concluye que la distancia por galón es una variable aleatoria que se distribuye normal con una desviación estándar conocida de 1.4 millas por galón. Con base en esta información, ¿se encuentra apoyada la sospecha del propietario con $\alpha = 0.01$? ¿Cuál es el valor p en este caso?

9.15. En el ejercicio 9.14, ¿cuántas veces debe observarse la distancia recorrida por galón para que con una probabilidad de 0.9 sea detectado un valor tan bajo como 28 mpg?

9.16. En cierto condado de Iowa, la cosecha promedio de maíz por acre fue de 100 toneladas por acre. Para un año dado en el que el clima fue particularmente bueno, se seleccionaron 12 parcelas en forma aleatoria y éstas arrojaron una cosecha promedio de 106 toneladas por acre, para la misma variedad de maíz. Si la producción por acre se modela en forma adecuada por una distribución normal con una desviación estándar de 8 toneladas por acre, ¿existe alguna razón para creer que este año la producción será mejor que la producción promedio normal? Empléese $\alpha = 0.01$. Para este caso, ¿cuál es el valor p?

9.17. Para el ejercicio 9.16, obtener el correspondiente intervalo inferior de confianza del 99% para el estimador del valor real promedio de la producción por acre, y deducir el intervalo de posibles valores para μ bajo la hipótesis nula para la que H_0 no puede rechazarse con el mismo valor de α.

9.18. En una planta de armado se diseña una operación específica la cual toma un tiempo promedio de 5 minutos. El gerente de la planta sospecha que para un operador en particular el tiempo promedio es diferente. El gerente toma una muestra de 11 tiempos de operación para este empleado y obtiene los siguientes resultados (en minutos): 4.8, 5.6, 5.3, 5.2, 4.9, 4.7, 5.7, 4.9, 5.7, 4.9, 4.6. Si se supone que el tiempo de operación se encuentra modelado en forma adecuada por una distribución normal:

a) ¿Se encuentra la sospecha del gerente apoyada por la evidencia con $\alpha = 0.02$? ¿cuál es el valor de p?
b) Obtener el correspondiente intervalo de confianza estimado del 99% para el tiempo promedio real, y deducir el intervalo de posibles valores de μ bajo H_0 para los que no puede rechazarse la hipótesis nula.

9.19. A veces los productos radioactivos de desecho industrial van a dar a las fuentes de agua que se utilizan para el consumo de la población. Por razones como ésta, las agencias estatales de salud vigilan en forma periódica las fuentes naturales de agua mediante la toma y el análisis de muestras de agua. En forma legal se ordena que la cantidad promedio de radiación en el agua para beber no debe exceder el valor de 4 picocuries por litro de agua. Se toma una muestra de 16 especímenes de una fuente natural de abasto de una zona densamente poblada, la cual proporciona valores para la media y la desviación estándar muestral de 4.2 y 1.2 picocuries por litro, respectivamente. Supóngase que la cantidad de radiación por litro de agua se encuentra modelada, en forma aproximada, por una distribución normal.

a) ¿Debe usarse un valor, en particular, pequeño para la probabilidad del error de tipo I en esta situación? ¿Por qué?

b) Selecciónese un valor de alfa y pruébense las hipótesis adecuadas. ¿Cuál es el valor de p?

c) ¿Debería preocupar la suposición de normalidad? Coméntese.

9.20. En el ejercicio 9.14, supóngase que la desviación estándar del rendimiento en distancia por galón no se conoce. Pruébese la misma hipótesis del ejercicio 9.14 y compárense los resultados.

9.21. En el ejercicio 9.11, supóngase que se asume un tiempo de falla el cual se encuentra normalmente distribuido. Pruébese la misma hipótesis del ejercicio 9.11 y compárense los resultados.

9.22. Considérese la prueba de H_0: $p = p_0$ contra H_1: $p = p_1$ para el parámetro binomial p, en donde $p_1 > p_0$. Mediante el empleo del lema de Neyman-Pearson, demuéstrese que la mejor región crítica de tamaño α se basa en el número de éxitos observados en los n ensayos independientes.

9.23. Un fabricante de lavadoras afirma que sólo el 5% de todas las unidades que vende sufren una falla durante el primer año de operación normal. Una organización de consumidores ha pedido a 20 familias de igual número de miembros que han adquirido éstas lavadoras, que reporten cualquier mal funcionamiento durante el primer año. Al final de éste, sólo tres familias reportaron mal funcionamiento.

a) Si la organización de consumidores cree que la proporción de lavadoras que sufrirán alguna falla es más alta que el valor afirmado por el fabricante, empléese el ejercicio 9.22 para determinar si puede rechazarse H_0: $p = 0.05$ con un tamaño máximo del error de tipo I de 0.1.

b) Mediante el empleo de un método aproximado basado en el material de la sección 8.4.6, pruébese la hipótesis nula y compárense las probabilidades de las estadísticas de prueba, asumiendo valores tan extremos o más de los determinados, dado que H_0 es cierta.

9.24. Supóngase que en una muestra aleatoria de 20 bebés concebidos mediante un proceso de fertilización *in vitro,* 15 son mujeres.

a) Mediante el uso del ejercicio 9.22, determínese qué tan probable es el tener 15 o más mujeres, si la verdadera proporción de éstas es de 0.5.

b) Compárese la probabilidad de la parte *a* con la que se obtiene mediante el empleo de la aproximación normal.

9.25. Una organización de salud se interesa en actualizar su información con respecto a la proporción de hombres que fuman. Con base en estudios previos, se cree que la proporción es del 40%. La organización lleva a cabo una encuesta en la que se seleccionan en forma aleatoria 1 200 hombres a los cuales se les preguntan sus hábitos de fumador. De los 1 200, 420 son fumadores. Emplee un método aproximado para determinar si esta evidencia apoya la noción de que la proporción de hombres que fuman es diferente del 40% para $\alpha = 0.01$.

9.26. El responsable de la campaña política del candidato A piensa en el ambiente de las últimas semanas previas a las elecciones. Él piensa que su candidato se encuentra en igual posición que su oponente, el candidato B, pero han ocurrido algunos reveses en forma reciente. El responsable lleva a cabo una encuesta en 1 500 ciudadanos.Si de los 1 500 720 indican una preferencia por el candidato A, ¿existe alguna razón para creer que el candidato A se encuentra en desventaja con relación al candidato B? Empléese $\alpha = 0.05$.
$\alpha = 0.05$.

9.27. un fabricante desea comparar la tensión promedio de su hilo con la de su más cercano competidor. Las tensiones de 100 hilos para cada marca se observaron bajo condiciones controladas. Las medias y desviaciones estándar de cada marca fueron las siguientes:

$$\bar{x}_1 = 110.8 \qquad \bar{x}_2 = 108.2$$
$$s_1 = 10.2 \qquad s_2 = 12.4.$$

Si se supone que el muestreo se llevó a cabo sobre dos poblaciones normales e independientes, ¿existe alguna razón para creer que hay una diferencia entre las tensiones promedio de ruptura de los dos hilos? Úsese $\alpha = 0.02$. ¿Cuál es el valor de p? (Sugerencia: la estadística dada por (8.41) en la que los estimados s_1^2 y s_2^2 reemplazan a las correspondientes varianzas poblacionales es aproximadamente $N(0, 1)$ para valores grandes de n_1 y n_2).

9.28. En el ejercicio 9.27, obtener las curvas de potencia y característica de operación.

9.29. Obtener una expresión equivalente a (9.14) para probar H_0: $\mu_X - \mu_Y = \delta_0$ contra H_1: $\mu_X - \mu_Y = \delta_1 < \delta_0$.

9.30. Se cree que el promedio verbal para el número de respuestas correctas para la prueba SAT para las mujeres es mayor que el de los hombres por más de diez puntos. Las muestras aleatorias para ambos sexos arrojaron los siguientes resultados:

Hombres	*Mujeres*
$n_1 = 125$	$n_2 = 100$
$\bar{x}_1 = 480$	$\bar{x}_2 = 460$
$s_1 = 60$	$s_2 = 52$

a) Si se muestrearon dos poblaciones independientes normales, ¿se encuentra la creencia apoyada por la evidencia muestral con $\alpha = 0.05$? ¿Cuál es el valor de p?
b) Supóngase que la verdadera diferencia es de 15 puntos. ¿Cuál es la potencia de la prueba anterior?

9.31. Mediante el empleo de los datos del ejercicio 8.32, determine si existen diferencias estadísticamente discernibles para la tensión de ruptura de los metales producidos por los dos procesos con $\alpha = 0.05$. ¿Cuál es el valor de p?

9.32. A finales de la década de los setenta se descubrió que la sustancia carcionogénica nitrosodimetilamina (NDMA) se formaba durante el secado de la malta verde, la cual se empleaba para fabricar cerveza. A principios de los ochenta se desarrolló un nuevo proceso para el secado de la malta, el cual minimizaba la formación de NDMA. Se tomaron muestras aleatorias de una cerveza doméstica que se fabricó empleando ambos procesos de secado, y se tomaron los niveles de NDMA en partes por billón. Se obtuvieron los siguientes resultados:

Proceso anterior	6	4	5	5	6	5	5	6	4	6	7	4
Proceso propuesto	2	1	2	2	1	0	3	2	1	0	1	3

si se supone que se muestrearon dos distribuciones normales independientes con varianzas iguales, ¿existe alguna razón para creer, a un nivel de $\alpha = 0.05$ que ha disminuido la cantidad promedio de NDMA en más de dos partes por billón con el empleo del nuevo proceso?

9.33. Se espera que dos operadores produzcan, en promedio, el mismo número de unidades terminadas en el mismo tiempo. Los siguientes datos son los números de unidades terminadas para ambos trabajadores en una semana de trabajo:

Operador 1	*Operador 2*
12	14
11	18
18	18
16	17
13	16

Si se supone que el número de unidades terminadas diariamente por los dos trabajadores son variables aleatorias independientes distribuidas normales con varianzas iguales, ¿se puede discernir alguna diferencia entre las medias a un nivel $\alpha = 0.1$?

9.34. En el ejercicio 9.33, dado que los datos son observaciones diarias sobre un periodo de una semana, ¿debe usted considerar un enfoque alternativo a este problema? Discuta las ventajas de este enfoque y demuestre que se obtienen resultados diferentes a los del ejercicio 9.33. ¿Por qué se obtienen resultados diferentes?

9.35. Un investigador médico se interesa en comparar la efectividad de dos dietas muy populares, A y B. En particular, el investigador desea determinar si una dieta es más efectiva para reducir el peso de las personas obesas en un lapso dado de tiempo. Discuta de manera completa el cómo debe el investigador llevar a cabo su experimento. Asegúrese de indicar las suposiciones necesarias.

9.36. Un educador ha desarrollado una nueva prueba de aptitud mucho más breve que la que se encuentra en uso. El educador desea comparar las dos pruebas. Discuta el enfoque que empleará el educador para hacer posible tal comparación.

9.37. Un fabricante desea comparar el proceso de armado común para uno de sus productos con un método propuesto que supuestamente reduce el tiempo de armado. Se seleccionaron ocho trabajadores de la planta de armado y se les pidió que armaran las unidades con ambos procesos. Los siguientes son los tiempos observados en minutos.

Trabajador	*Proceso actual*	*Proceso propuesto*
1	38	30
2	32	32
3	41	34
4	35	37
5	42	35
6	32	26
7	45	38
8	37	32

a) En $\alpha = 0.05$ ¿existe alguna razón para creer que el tiempo de armado para el proceso actual es mayor que el del método propuesto por más de dos minutos?

b) ¿Qué suposiciones son necesarias para probar la hipótesis del inciso *a,* y cuál es el valor de *p*?

c) Obténgase un intervalo de confianza del 95% para la diferencia entre las medias de los tiempos de armado.

9.38. Se llevó a cabo un estudio para determinar el grado en el cual el alcohol entorpece la habilidad de pensamiento para llevar a cabo determinada tarea. Se seleccionaron al azar diez personas de distintas características y se les pidió que participaran en el experimento. Después de proporcionarles la información pertinente, cada persona llevó a cabo la tarea sin nada de alcohol en su organismo. Entonces, la tarea volvió a llevarse a cabo, después de que cada persona había consumido una cantidad suficiente de alcohol para tener un contenido en su organismo de 0.1%.

a) Discutir los aspectos importantes de control que el experimentador debe considerar al llevar a cabo el experimento.

b) Supóngase que los tiempos "antes" y "después" (en minutos) de los diez participantes son los siguientes:

Participante	*Antes*	*Después*
1	28	39
2	22	45
3	55	67
4	45	61
5	32	46
6	35	58
7	40	51
8	25	34
9	37	48
10	20	30

¿Puede concluirse a un nivel de $\alpha = 0.05$ que el tiempo promedio "antes" es menor que el tiempo promedio "después" por más de diez minutos?

9.39. En el ejercicio 9.19, ¿existe alguna razón para creer que la varianza en la cantidad de radiación en la fuente de agua es mayor de 1.25 picocuries cuadrados? Emplee $\alpha = 0.05$.

9.40. Desarróllense expresiones generales para calcular la probabilidad del error de tipo II cuando se prueban las hipótesis H_0: $\sigma^2 = \sigma_0^2$ contra cualquiera de las dos siguientes alternativas H_1: $\sigma^2 > \sigma_0^2$ y H_1: $\sigma^2 < \sigma_0^2$.

9.41. Empléense los resultados del ejercicio 9.40 para obtener la potencia de la prueba de la hipótesis en el ejercicio 9.39 si $\sigma^2 = 1.4$.

9.42. El gerente de una planta sospecha que el número de piezas que produce un trabajador en particular por día, fluctúa más allá del valor normal esperado. El gerente decide observar el número de piezas que produce este trabajador durante diez días, seleccionados éstos al azar. Los resultados son 15, 12, 8, 13, 12, 15, 16, 9, 8, y 14. Si se sabe que la desviación estándar para todos los trabajadores es de dos unidades y si el número de éstas que se produce diariamente, se encuentra modelado en forma adecuada por una distribución normal, a un nivel de $\alpha = 0.05$, ¿tiene apoyo la sospecha del gerente? ¿Cuál es el valor de p?

9.43. En un proceso de llenado, la tolerancia para el peso de los recipientes es de ocho gramos. Para reunir este requisito, la desviación estándar en el peso debe ser de dos gramos. Los pesos de 25 recipientes seleccionados al azar dieron como resultado una desviación estándar de 2.8 gramos.

a) Si los pesos se encuentran normalmente distribuidos, determinar si la varianza de éstos es diferente del valor necesario. Empléese $\alpha = 0.02$.

b) ¿Para qué valores de la varianza muestral no puede rechazarse la hipótesis nula del inciso *a*? ¿Se encuentran estos valores equidistantes del valor necesario de la varianza? ¿Cómo deberían ser? Coméntese.

9.44. Considérense los datos del ejercicio 9.32. Para un nivel de $\alpha = 0.05$ ¿existe alguna razón para pensar que las varianzas no son iguales?

9.45. Un inversionista desea comparar los riesgos asociados con dos diferentes mercados, A y B. El riesgo de un mercado dado se mide por la variación en los cambios diarios de precios. El inversionista piensa que el riesgo asociado con el mercado B es mayor que el del mercado A. Se obtienen muestras aleatorias de 21 cambios de precio diarios para el mercado A y de 16 para el mercado B. Se obtienen los siguientes resultados:

Mercado A	*Mercado B*
$\bar{x}_A = 0.3$	$\bar{x}_B = 0.4$
$s_A = 0.25$	$s_B = 0.45$

a) Si se supone que las muestras provienen de dos poblaciones normales e independientes a un nivel de $\alpha = 0.05$ ¿encuentra apoyo la creencia del inversionista?
b) Si la varianza muestral de A es la dada, ¿cuál es el máximo valor de la varianza muestral de B con base en $n = 16$ que no llevará al rechazo de la hipótesis nula del inciso *a*?

9.46. Para el ejercicio 9.33, ¿puede apoyarse la opinión de que la variación en el número de artículos terminados para el operador 2 es menor que para el operador 1 a un nivel $\alpha = 0.05$?

9.47. En un estudio reciente que abarcó 25 años, se investigó la posible protección que proporciona la ingestión de una forma de vitamina A llamada caroteno contra el desarrollo del cáncer pulmonar. Se encontró que de 488 hombres que habían ingerido una baja cantidad de esta sustancia durante este tiempo, 14 desarrollaron cáncer pulmonar, pero en un grupo del mismo tamaño en el que el consumo de caroteno era mayor, sólo dos personas desarrollaron cáncer. Bajo las suposiciones apropiadas, ¿puede concluirse que la ingestión de caroteno reduce el riesgo de desarrollar cáncer pulmonar en los hombres? Empléese $\alpha = 0.01$. ¿Cuál es el valor de p? Desde un punto de vista estadístico, ¿qué consejo se podría dar al investigador médico que se interesa en un proyecto como éste?

9.48. Para el ejercicio 9.47, determinar un intervalo de confianza estimado del 99% para la verdadera diferencia entre las dos proporciones.

9.49. Un economista al servicio de una agencia estatal desea determinar si la frecuencia de desempleo en dos grandes áreas urbanas del estado son diferentes. Con base en muestras aleatorias de cada ciudad, cada una de 500 personas, el economista encuentra 35 personas desempleadas en un área y 25 en la otra. Bajo las suposiciones adecuadas y con un nivel $\alpha = 0.05$ ¿existe alguna razón para creer que las frecuencias de desempleo en las dos áreas son diferentes? ¿Cuál es el valor de p?

9.50. Un usuario de grandes cantidades de componentes aléctricos adquiere éstos principalmente de dos proveedores, A y B. Debido a una mejor estructura en precios, el usuario hará negocio únicamente con el proveedor B si la proporción de artículos defectuosos para A y para B es la misma. De dos grandes lotes, el usuario selecciona al azar 125 unidades de A y 100 unidades de B; inspecciona las unidades y encuentra siete y siete unidades defectuosas, respectivamente. Bajo las suposiciones adecuadas y con base en esta información, ¿existe alguna razón para no comprar en forma única las componentes del proveedor B? Empléese $\alpha = 0.02$.

CAPÍTULO DIEZ

Pruebas de bondad de ajuste y análisis de tablas de contingencia

10.1 Introducción

Recuérdese que una hipótesis estadística es una afirmación con respecto a una característica que se desconoce de una población de interés. En el capítulo 9 fue, en forma exclusiva, el valor de algún parámetro θ. En este capítulo se examinarán las pruebas de hipótesis estadísticas en las que la característica que se desconoce es alguna propiedad de la forma funcional de la distribución que se muestrea. Además, se discutirán las pruebas de independiencia entre dos variables aleatorias en las cuales la evidencia muestral se obtiene mediante la clasificación de cada variable aleatoria en un cierto número de categorías.

En forma tradicional, este tipo de prueba recibe el nombre de *bondad del ajuste* ya que ésta compara los resultados de una muestra aleatoria con aquéllos que se espera observar si la hipótesis nula es correcta. La comparación se hace mediante la clasificación de los datos que se observan en cierto número de categorías y entonces comparando las frecuencias observadas con las esperadas para cada categoría. Para un tamaño específico del error de tipo I, la hipótesis nula será rechazada si existe una diferencia suficiente entre las frecuencias observadas y las esperadas.

Vale la pena notar que para situaciones de este tipo la hipótesis alternativa es compuesta y, en muchas ocasiones, no se encuentra identificada en forma explícita. El resultado es que la función de potencia es muy difícil de obtener en forma analítica. En consecuencia, una prueba de bondad de ajuste no debe usarse por sí misma para aceptar la afirmación de la hipótesis nula. La decisión es no rechazar H_0 (más que aceptarla) si la diferencia que existe entre las frecuencias observadas y esperadas es, en forma relativa, pequeña.

10.2 La prueba de bondad de ajuste chi-cuadrada

Una prueba de bondad de ajuste se emplea para decidir cuándo un conjunto de datos se apega a una distribución de probabilidad dada. Considérese una muestra aleatoria de tamaño n de la distribución de una variable aleatoria X dividida en k clases exhaustivas y mutuamente excluyentes, y sea N_i, $i = 1, 2, \ldots, k$, el número de observaciones en la i-ésima clase. Considérese la verificación de la hipótesis nula

$$H_0: F(x) = F_0(x), \tag{10.1}$$

en donde el modelo de probabilidad propuesto $F_0(x)$ se encuentra especificado, de manera completa, con respecto a todos los parámetros. De esta forma la hipótesis nula es sencilla. Dado que se especifica $F_0(x)$ de manera completa, se puede obtener la probabilidad p_i de obtener una observación en la i-ésima clase bajo H_0, en donde necesariamente $\sum_{i=1}^{k} p_i = 1$.

Sea n_i la realización de N_i para $i = 1, 2 \ldots k$ de manera tal que $\sum_{i=1}^{k} n_i = n$. La probabilidad de tener, de manera exacta, n_i observaciones en la i-ésima clase es $p_i^{n_i}$ para $i = 1, 2 \ldots k$. Dado que existen k categorías mutuamente excluyentes con probabilidades $p_1, p_2, \ldots, p_k$, entonces bajo la hipótesis nula la probabilidad de la muestra agrupada es igual a la función de probabilidad de una distribución multinomial determinada (6.3).

Para deducir una prueba estadística adecuada para H_0, considérese el caso en el que $k = 2$. Este es la distribución binomial con una función de probabilidad dada por (4.1) y en la que $x = n_1$, $p = p_1$, $n - x = n_2$, y $1 - p = p_2$. Considérese la variable aleatoria estandarizada

$$Y = \frac{N_1 - np_1}{\sqrt{np_1(1 - p_1)}}.$$

Del capítulo 5, recuérdese que para un valor de n suficientemente grande, la distribución de Y es aproximadamente igual a la normal estándar. Además, del ejemplo 5.14 se sabe que el cuadrado de una variable aleatoria normal estándar tiene una distribución chi-cuadrada con un grado de libertad. Entonces, la estadística

$$\begin{aligned}
\frac{(N_1 - np_1)^2}{np_1(1 - p_1)} &= \frac{(N_1 - np_1)^2}{np_1} + \frac{(N_1 - np_1)^2}{np_2} \\
&= \frac{(N_1 - np_1)^2}{np_1} + \frac{[n - N_2 - n(1 - p_2)]^2}{np_2} \\
&= \frac{(N_1 - np_1)^2}{np_1} + \frac{(N_2 - np_2)^2}{np_2} \\
&= \sum_{i=1}^{2} \frac{(N_i - np_i)^2}{np_i}
\end{aligned}$$

tiene aproximadamente una distribución chi-cuadrada con un grado de libertad conforme n va tomando valores cada vez más grandes.

Si se sigue este tipo de razonamiento, puede demostrarse que para $k \geq 2$ categorías distintas, la estadística

$$\sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} \tag{10.2}$$

tiene una distribución, en forma aproximada, chi-cuadrada con $k - 1$ grados de libertad, si n tiene un valor suficientemente grande. Nótese que N_i es la frecuencia observada en la i-ésima clase, y np_i es la frecuencia correspondiente que se esperaba bajo la hipótesis nula. De acuerdo con lo anterior, la estadística es la suma sobre todas las k clases de los cocientes de los cuadrados de las diferencias entre las frecuencias observada y esperada, y la frecuencia esperada. La estadística dada por (10.2) recibe el nombre de ***prueba de bondad de ajuste chi-cuadrada*** de Pearson. Si existe una concordancia perfecta entre las frecuencias que se observaban y las que se esperaban, la estadística tendrá un valor igual a cero; por otro lado, si existe gran discrepancia entre estas frecuencias, la estadística tomará un valor muy grande. Por ello se desprende que para un tamaño dado del error de tipo I, la región crítica es el extremo superior de una distribución chi-cuadrada con $k - 1$ grados de libertad.

Ejemplo 10.1 El gerente de una planta industrial pretende determinar si el número de empleados que asisten al consultorio médico de la planta se encuentra distribuido, en forma equitativa, durante los cinco días de trabajo de la semana. Con base en una muestra aleatoria de cuatro semanas completas de trabajo, se observó el siguiente número de consultas:

Lunes	*Martes*	*Miércoles*	*Jueves*	*Viernes*
49	35	32	39	45

Con $\alpha = 0.05$, ¿existe alguna razón para creer que el número de empleados que asisten al consultorio médico, no se encuentra distribuido en forma equitativa durante los días de trabajo de la semana?

Una distribución uniforme implicaría que las proporciones para cada día de la semana sean iguales. Por lo tanto, deberá probarse la hipótesis nula

$$H_0\colon p_i = 0.2, \qquad i = 1, 2, \ldots, 5.$$

Dado que el tamaño de la muestra es $n = 200$, la frecuencia esperada para cada día es $np_i = 40$. Entonces, el valor de la estadística de prueba es

$$\chi^2 = \frac{(49-40)^2}{40} + \frac{(35-40)^2}{40} + \frac{(32-40)^2}{40} + \frac{(39-40)^2}{40} + \frac{(45-40)^2}{40} = 4.9.$$

Para $k = 5$ clases, se observa que el valor crítico es $\chi^2_{0.95,\,4} = 9.49$. Ya que $\chi^2 = 4.9 < \chi^2_{0.95,\,4} = 9.49$, no puede rechazarse la hipótesis nula. Con base en esta evidencia, no existe ninguna razón para creer que el número de empleados que acuden al

consultorio no se encuentre distribuido en forma uniforme a lo largo de la semana de trabajo.

Una ventaja de la prueba de bondad de ajuste chi-cuadrada es que para valores grandes de n, la distribución límite chi-cuadrada de la estadística, es independiente a la forma de la distribución propuesta $F_0(x)$ bajo H_0. Como resultado se tiene que la prueba de bondad de ajuste chi-cuadrada también se emplea en situaciones en las que $F_0(x)$ es continua. Sin embargo, debe hacerse énfasis en que la naturaleza de la prueba de bondad de ajuste chi-cuadrada es discreta en el sentido en el que ésta compara las frecuencias que se observan y se esperan para un número finito de categorías. De acuerdo con lo anterior, si $F_0(x)$ es continua, la prueba no compara las frecuencias que se observan alisadas con la función de densidad propuesta tal como lo implica la hipótesis nula. Más bien, la comparación se lleva a cabo aproximando la distribución continua bajo H_0 con un número finito de intervalo de clase. A pesar de esta limitación, la prueba de bondad de ajuste chi-cuadrada es un procedimiento razonablemente adecuado para probar suposiciones de normalidad siempre y cuando el tamaño de la muestra sea, en forma moderada, grande. Con respecto a la pregunta de qué tan grande debe ser el tamaño de la muestra, se ha encontrado que con n igual a cinco veces el número de clases, los resultados son aceptables. Una regla conservadora a seguir es el seleccionar un muestra de manera tal que toda frecuencia esperada no sea menor que cinco. Lo anterior puede lograrse combinando clases vecinas pero, para cada par de clases que se combina, el número de grados de libertad debe reducirse en uno.

A menos que pueda especificarse una hipótesis alternativa que consista en un modelo alternativo $F_1(x)$ particular, la potencia de la prueba de bondad de ajuste chi-cuadrada es muy difícil de determinar en forma analítica. Sin embargo, puede demostrarse que la potencia tiende a 1 conforme n tiende a ∞. Este resultado implica que para muestras de gran tamaño es casi seguro el rechazar la hipótesis nula debido a que es muy difícil especificar una H_0 lo suficientemente cercana a la verdadera distribución. De esta forma, la aplicabilidad de la prueba de bondad de ajuste chi-cuadrada es cuestionable cuando se tienen muestras de tamaño muy grande.

Ejemplo 10.2 En la tabla 5.2 se proporcionan los datos que se agrupan para el número de respuestas correctas para la prueba SAT de matemáticas, de los alumnos del tercer año de preparatoria. Recuérdese que en el ejemplo 5.5 se compararon las frecuencias que se observaron con las que se esperaron, en donde estas últimas se obtuvieron con base en una distribución normal con media 491 y desviación estándar igual a 120. Con base en la prueba de bondad de ajuste chi-cuadrada, ¿existe alguna razón para creer que el número de respuestas correctas para la prueba de matemáticas SAT no se encuentran distribuidas normalmente con media 491 y desviación estándar igual a 120 a un nivel de $\alpha = 0.01$?

Considérese la prueba de la siguiente hipótesis nula

$$H_0: F(x) = F_0(x),$$

en donde $F_0(x)$ es el modelo de probabilidad normal con media 491 y desviación estándar 120. Bajo la hipótesis nula, las frecuencias esperadas para las 12 clases se

encuentran en la última columna de la tabla 5.2. Éstas se determinaron primero convirtiendo cada intervalo de cada clase al correspondiente intervalo normal estándar, empleando para esto $\mu = 491$ y $\sigma = 120$. Después se determinó la probabilidad de cada intervalo bajo H_0. Finalmente, para cada clase, el valor de probabilidad se multiplicó por el tamaño de la muestra $n = 478\ 193$ para obtener la frecuencia esperada. Nótese que las probabilidades que aparecen en la penúltima columna de la tabla 5.2 no suman uno. Pero bajo la hipótesis nula las clases deben ser exhaustivas, de manera tal que $\Sigma_{i=1}^{k} p_i = 1$. Lo anterior puede lograrse mediante el ajuste de las clases primera y última de manera tal que la primera no tenga límite inferior y la última no tenga límite superior. Dado que bajo H_0, $X \sim N(491, 120)$,

$$P(X \leqslant 250) = P(Z \leqslant -2.01) = 0.0222,$$

y la frecuencia modificada para la primera clase es (478 193) (0.0222) = 10 615.88. De manera similar para la última clase

$$P(X \geqslant 750) = P(Z \geqslant 2.16) = 0.0154,$$

lo cual da como resultado una frecuencia esperada de 7 364.17.

Con base en las 12 clases, el valor de la estadística chi-cuadrada es

$$\chi^2 = \frac{(3\ 423 - 10\ 615.88)^2}{10\ 615.88} + \frac{(18\ 434 - 16\ 115.10)^2}{16\ 115.10} + \cdots + \frac{(6\ 414 - 7\ 364.17)^2}{7\ 364.17}$$
$$= 13\ 067.02,$$

el cual se encuentra, en forma clara, más allá del valor crítico $\chi^2_{0.99,\,11} = 24.75$. De acuerdo con lo anterior, la hipótesis nula de que el número de respuestas correctas para la prueba SAT se encuentra normalmente distribuido con media 491 y desviación estándar de 120, debe rechazarse. Este ejemplo ilustra el comentario formulado con anterioridad con respecto a muestras de gran tamaño, en donde la hipótesis nula casi seguramente resulta rechazada.

Recuérdese que la hipótesis nula dada por (10.1) es simple ya que el modelo de probabilidad propuesto $F_0(x)$ se especificó de manera completa con respecto a todos sus parámetros. Sin embargo, para muchas aplicaciones que toman en cuenta la bondad del ajuste, sólo puede especificarse la forma de $F_0(x)$. Por ejemplo, supóngase que se desea probar la hipótesis nula de que un conjunto de observaciones de una medida de interés X se ajustan a una distribución normal, pero no puede especificarse el valor de la media o el de la variaza. Lo anterior da como resultado que la hipótesis nula

$$H_0\text{: } F(x) = F_0(x)$$

es compuesta. En consecuencia, se tiene que las frecuencias esperadas np_i para las $i = 1, 2 \ldots k$ clases no pueden determinarse, ya que éstas son funciones de los parámetros desconocidos de $F_0(x)$.

Supóngase que T es una estadística para un parámetro desconocido θ de $F_0(x)$. En el contexto de la prueba de bondad de ajuste, tanto las frecuencias observables

N_i como las frecuencias esperadas $np_i(T)$ son variables aleatorias, en donde $p_i(T)$ indica que las probabilidades bajo la hipótesis nula son funciones de la estadística T de θ. Puede demostrarse que si para cualquier parámetro desconocido θ la estadística T es el estimador de máxima verosimilitud de θ, y si las frecuencias esperadas se determinan como funciones de los estimadores de máxima verosimilitud, entonces

$$\sum_{i=1}^{k} \frac{[N_i - np_i(T)]^2}{np_i(T)} \tag{10.3}$$

tiene aproximadamente una distribución chi-cuadrada con $k - 1 - r$ grados de libertad, para valores de n grandes, en donde r es el número de parámetros que se está tratando de estimar.

Al igual que en el caso previo en el que se tenía una H_0, sencilla, la región crítica es el extremo superior de la distribución chi-cuadrada. Pero, a diferencia del caso anterior, el número de grados de libertad se reduce por una cantidad igual al número de parámetros que se están estimando. Como consecuencia, existe un corrimiento hacia la izquierda en el valor crítico para el mismo tamaño del error de tipo I, y la hipótesis nula puede rechazarse para un valor observado más pequeño de (10.3) que en el caso previo. Lo anterior es lógico ya que el ajuste deberá ser mejor debido a que los parámetros desconocidos se estiman con base en las observaciones de la muestra.

Las características importantes para la aplicación de la prueba de bondad de ajuste chi-cuadrada para el caso compuesto son idénticas a las que tienen para la hipótesis nula simple. Surge un problema relativamente pequeño al decidir si los parámetros desconocidos deberán estimarse con base en los datos que se agruparon en los que no. En forma teórica, ninguno de los dos enfoques puede ser el correcto debido a que los estimados de máxima verosimilitud deben obtenerse maximizando la función de verosimilitud con base en la distribución multinomial. En forma afortunada, resulta que la mayoría de las veces el error que se comete no es serio. De esta forma, se pueden utilizar los estimados de máxima verosimilitud obtenidos, ya sea de los datos agrupados o de los no agrupados, en forma segura.

Ejemplo 10.3 Recuérdese el ejemplo 4.5 en el que se compararon el número de anotaciones de seis puntos por equipo y por juego en la NFL con el número que esperaban de éstos, si el número de anotaciones de seis puntos tiene una distribución de Poisson. Con base en la información contenida en la tabla 4.3, ¿existe alguna razón para creer, a un nivel de 0.05, que el número de anotaciones no es variable aleatoria de Poisson?

Dado que el valor del parámetro de Poisson λ no se especifica, el estimado de máxima verosimilitud de λ con base en la información que se proporcionó en la tabla 4.3 es $\hat{\lambda} = 2.435$. Bajo la hipótesis nula de una distribución de Poisson, la probabilidad de tener cero anotaciones es

$$p(0) = (2.435)^0 \exp(-2.435)/0! = 0.0876.$$

Para $n = 448$, el número esperado de cero anotaciones es (448)(0.0876) = 39.24. Si se sigue este procedimiento, pueden obtenerse las demás frecuencias esperadas. En la tabla 10.1, se presenta el cálculo de la estadística chi-cuadrada.

TABLA 10.1 Cálculo de la estadística chi-cuadrada para el ejemplo 10.3

Número de anotaciones	*Frecuencia observada*	*Frecuencia esperada*	$\frac{[n_i - np_i(\hat{\lambda})]^2}{np_i(\hat{\lambda})}$
0	35	39.24	0.458
1	99	95.56	0.124
2	104	116.34	1.309
3	110	94.44	2.564
4	62	57.48	0.355
5	25	28.00	0.321
6	10	11.38	0.167
7	3	5.56	1.179
Totales	448	448	6.477

Para $k = 8$ categorías y con un parámetro estimado, el número de grados de libertad es 6. Para $\alpha = 0.05$ el valor crítico es $\chi^2_{0.95,6} = 12.60$. Dado que $\chi^2 = 6.477 < \chi^2_{0.95,6} = 12.60$, no puede rechazarse la hipótesis nula de que el número de anotaciones de seis puntos por equipo en la NFL es una variable aleatoria de Poisson.

10.3 La estadística de Kolmogorov-Smirnov

Recuérdese que para aplicar la prueba de bondad de ajuste chi-cuadrada cuando el modelo propuesto bajo H_0 es continuo, es necesario aproximar $F_0(x)$ mediante el agrupamiento de los datos observados en un número finito de intervalos de clase. Este requisito de agrupar los datos implica tener una muestra de tamaño más o menos grande. De esta manera, la prueba de bondad de ajuste chi-cuadrada se encuentra limitada cuando $F_0(x)$ es continua y la muestra aleatoria disponible tiene un tamaño pequeño. Una prueba de bondad de ajuste más apropiada que la chi-cuadrada cuando $F_0(x)$ es continua, es la basada en la estadística de Kolmogorov-Smirnov. La prueba de Kolmogorov-Smirnov no necesita que los datos se encuentren agrupados y es aplicable a muestras de tamaño pequeño. Ésta se basa en una comparación entre las funciones de distribución acumulativa que se observan en la muestra ordenada y la distribución propuesta bajo la hipótesis nula. Si esta comparación revela una diferencia suficientemente grande entre las funciones de distribución muestral y propuesta, entonces la hipótesis nula de que la distribución es $F_0(x)$, se rechaza.

Considérese la hipótesis nula por (10.1), en donde $F_0(x)$ se especifica en forma completa. Denótense por $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ a las observaciones ordenadas de una muestra aleatoria de tamaño n y defínase la función de distribución acumulativa muestral como

$$S_n(x) = \begin{cases} 0 & x < x_{(1)}, \\ k/n & x_{(k)} \leqslant x < x_{(k+1)}, \\ 1 & x \geqslant x_n. \end{cases} \tag{10.4}$$

En otras palabras, para cualquier valor ordenado x de la muestra aleatoria, $S_n(x)$ es la proporción del número de valores en la muestra que son iguales o menores a x. Ya que $F_0(x)$ se encuentra completamente especificada, es posible evaluar a $F_0(x)$ para algún valor deseado de x, y entonces comparar este último con el valor correspondiente de $S_n(x)$. Si la hipótesis nula es verdadera, entonces es lógico esperar que la diferencia sea relativamente pequeña. La estadística de Kolmogorov-Smirnov se define como

$$D_n = \max_x | S_n(x) - F_0(x) |. \tag{10.5}$$

La estadística D_n tiene una distribución que es independiente del modelo propuesto bajo la hipótesis nula. Por esta razón, se dice D_n es una estadística independiente de la distribución. Lo anterior da como resultado que la función de distribución de D_n pueda evaluarse sólo en función del tamaño de la muestra y después usarse para cualquier $F_0(x)$. En la tabla J del apéndice, se proporcionan los valores cuantiles superiores de D_n para varios tamaños de la muestra. El lector debe notar que los valores asintóticos de d_n que se encuentran en la parte inferior de la tabla proporcionan una adecuada aproximación para valores de n mayores de 50.

Para un tamaño α del error de tipo I, la región crítica es de la forma

$$P\left(D_n > \frac{c}{\sqrt{n}}\right) = \alpha.$$

De acuerdo con lo anterior, la hipótesis H_0 se rechaza si para algún valor x observado el valor de D_n se encuentra dentro de la región crítica de tamaño α.

Como se hizo notar anteriormente, la estadística de Kolmogorov-Smirnov es, en general, superior a la prueba de bondad de ajuste chi-cuadrada cuando los datos involucran una variable aleatoria continua, debido a que no es necesario agrupar los datos. Además, la prueba de Kolmogorov-Smirnov tiene la atractiva propiedad de ser aplicable a muestras de tamaño pequeño. Por otro lado, la estadística se encuentra limitada, ya que el modelo propuesto bajo H_0 debe especificarse en forma completa. La estadística de Kolmogorov-Smirnov no se aplica a todos aquellos casos para los que las observaciones no son inherentemente cuantitativas a consecuencia de las ambigüedades que pueden surgir cuando se ordenan las observaciones.

Ejemplo 10.4 A continuación se proporcionan los valores ordenados de una muestra aleatoria del número de respuestas correctas para la SAT que se aplicó a todos los estudiantes que ingresaron a una universidad: 852, 875, 910, 933, 957, 963, 981, 998, 1007. 1010, 1015, 1018, 1023, 1035, 1048, 1063. En años anteriores, el número de respuestas correctas estaba representado, en forma adecuada, por una distribución normal con media 985 y desviación estándar 50. Con base en esta muestra, ¿existe alguna razón para creer que ha ocurrido un cambio en la distribución de respuestas correctas para la prueba SAT en esta universidad? Empléese un nivel $\alpha = 0.05$.

Sea X la variable aleatoria que representa el número de respuestas correctas para la prueba SAT. Considérese la prueba de la siguiente hipótesis nula

$$H_0\colon F(x) = F_0(x),$$

donde $F_0(x)$ es la función de distribución normal con media 985 y desviación estándar 50. Dado que X es una variable aleatoria continua y el tamaño de la muestra de X es pequeño, se usará la estadística de Kolmogorov-Smirnov para probar a H_0. La función de distribución muestral se obtiene mediante el empleo de (10.4) para los valores ordenados. Lo anterior involucra un incremento de $1/6 = 0.0625$ al valor previo de la distribución muestral. Los valores correspondientes del modelo normal propuesto se obtienen estandarizando primero a $N(0, 1)$ y empleando la tabla D del apéndice. En la tabla 10.2 se encuentra la información más importante.

Se observa que la máxima desviación es de 0.1207. De la tabla J del apéndice, el valor crítico de D_{16} para $\alpha = 0.05$ es 0.328. Dado que $0.1207 < 0.328$, no puede rechazarse la hipótesis nula. De acuerdo con ello no es posible detectar un cambio en la distribución para el número de respuestas correctas de la prueba SAT de la ya establecida $N(985, 50)$.

10.4 La prueba chi-cuadrada para el análisis de tablas de contingencia con dos criterios de clasificación

Muchas veces surge la necesidad de determinar si existe alguna relación entre dos rasgos diferentes en los que una población ha sido clasificada y en donde cada rasgo se encuentra subdividido en cierto número de categorías. Por ejemplo, ¿existe una relación entre el fumar cigarrillos y la predisposición a desarrollar cáncer pulmonar?, o también ¿existe una relación entre la filiación política y la opinión con respecto a incrementar el presupuesto armamentista? En ambos ejemplos, se ha clasificado a la población en dos características y en donde se supone que cada una de

TABLA 10.2 Cálculo de la estadística de Kolmogorov-Smirnov para el ejemplo 10.4

Valores ordenados	$S_n(x)$	$F_0(x)$	$\lvert S_n(x) - F_0(x)\rvert$
852	0.0625	0.0039	0.0586
875	0.1250	0.0139	0.1111
910	0.1875	0.0668	0.1207
933	0.2500	0.1492	0.1008
957	0.3125	0.2877	0.0248
963	0.3750	0.3300	0.0450
981	0.4375	0.4681	0.0306
998	0.5000	0.6026	0.1026
1007	0.5625	0.6700	0.1075
1010	0.6250	0.6915	0.0665
1015	0.6875	0.7257	0.0382
1018	0.7500	0.7454	0.0046
1023	0.8125	0.7764	0.0361
1035	0.8750	0.8413	0.0337
1048	0.9375	0.8962	0.0413
1063	1.0000	0.9406	0.0594

éstas tiene por lo menos dos categorías exhaustivas y mutuamente excluyentes. En el primer ejemplo las dos características son, si se es fumador, y si desarrolla cáncer pulmonar. Las categorías para estas dos características podrían ser si se es fumador crónico, moderado o no fumador, para la primera, y el si se desarrolla o no cáncer pulmonar para la segunda.

Cuando una muestra aleatoria que se obtiene de una población se clasifica de esta manera, el resultado recibe el nombre de ***tabla de contingencia con dos criterios de clasificación***. Esta tabla se forma por las frecuencias relativas que se observaron para las dos clasificaciones y sus correspondientes categorías. A pesar de que sólo se analizarán tablas de contingencia con dos clasificaciones, es posible analizar tablas que contengan más de dos clasificaciones.

El análisis de una tabla de este tipo supone que las dos clasificaciones son independientes. Esto es, bajo la hipótesis nula de independencia se desea saber si existe una diferencia suficiente entre las frecuencias que se observan y las correspondientes frecuencias que se esperan, tal que la hipótesis nula se rechace. La prueba chi-cuadrada, discutida en la sección 10.2, proporciona los medios apropiados para analizar este tipo de tablas.

Sea n una muestra aleatoria de una población que se clasifica de acuerdo con dos características A y B, cada una de las cuales contiene un número r y c de categorías, respectivamente. Además, sea N_{ij} el número de observaciones en la categoría (i, j), de las características A y B, respectivamente, para $i = 1, 2 \dots r$ y $j = 1, 2 \dots c$. Entonces una tabla de contingencia es un arreglo matricial de $r \times c$, dado en la tabla 10.3, en donde las entradas representan las realizaciones de las variables aleatorias N_{ij}.

Nótese que el total del i-ésimo renglón es la frecuencia de la i-ésima categoría de característica A, sumando sobre todas las categorías de la característica B. De manera similar, el total de la j-ésima columna es la frecuencia observada de la j-ésima categoría de B sumada sobre todas las categorías de A. Sean

$$n_{i\cdot} = \sum_{j=1}^{c} n_{ij} \qquad i = 1, 2, \dots, r,$$

$$n_{\cdot j} = \sum_{i=1}^{r} n_{ij} \qquad j = 1, 2, \dots, c,$$

TABLA 10.3 Tabla de contingencia con dos clasificaciones

		Característica B				
	Categorías	1	2	⋯	c	*Totales*
Característica A	1	n_{11}	n_{12}	⋯	n_{1c}	$n_{1\cdot}$
	2	n_{21}	n_{22}	⋯	n_{2c}	$n_{2\cdot}$
	·	·	·	⋯	·	·
	·	·	·	⋯	·	·
	·	·	·	⋯	·	·
	r	n_{r1}	n_{r2}		n_{rc}	$n_{r\cdot}$
	Totales	$n_{\cdot 1}$	$n_{\cdot 2}$	⋯	$n_{\cdot c}$	n

los símbolos para denotar las sumas de los renglones y de las columnas, respectivamente, en donde la notación "punto" indica el subscripto sobre el cual se lleva a cabo la sumatoria.

Sea p_{ij} la probabilidad de que un objeto seleccionado al azar de una población de interés se encuentre en la categoría (i, j) de la tabla de contingencia. Sea $p_{i\cdot}$ la probabilidad (marginal) de que un objeto se encuentre en la categoría i de la característica A, y sea $p_{\cdot j}$ la probabilidad de que un objeto se encuentre en la categoría j de la característica B. Si las dos características son independientes, la probabilidad conjunta debe ser igual al producto de las probabilidades marginales. De esta forma puede establecerse la hipótesis nula de la siguiente manera:

$$H_0\colon p_{ij} = p_{i\cdot}p_{\cdot j} \qquad i = 1, 2, \ldots, r;\ j = 1, 2, \ldots, c. \tag{10.6}$$

Si pueden especificarse las probabilidades marginales $p_{i\cdot}$ y $p_{\cdot j}$, entonces, bajo la hipótesis nula, la estadística

$$\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(N_{ij} - np_{i\cdot}p_{\cdot j})^2}{np_{i\cdot}p_{\cdot j}} \tag{10.7}$$

tiene en forma aproximada una distribución chi-cuadrada con $rc - 1$ grados de libertad para valores grandes de n. Sin embargo, la mayoría de las veces pueden no conocerse los valores de las probabilidades marginales y, de esta forma, se estiman con base en la muestra. Afortunadamente, la prueba de bondad de ajuste chi-cuadrada permanece como la estadística apropiada para probar (10.6), siempre que se empleen los estimados de máxima verosimilitud y se reste un grado de libertad del total para cada parámetro que se esté estimando. Dado que $\sum_{i=1}^{r} p_{i\cdot} = 1$ y $\sum_{j=1}^{c} p_{\cdot j} = 1$, existen $r - 1$ parámetros de renglón y $c - 1$ de columna a ser estimados. De esta forma, el número de grados de libertad será $rc - 1 - (r - 1) - (c - 1) = rc - r - c + 1 = (r - 1)(c - 1)$.

Puede demostrarse que los estimados de máxima verosimilitud de $p_{i\cdot}$ y $p_{\cdot j}$ están dados por

$$\hat{p}_{i\cdot} = n_{i\cdot}/n, \tag{10.8}$$

y

$$\hat{p}_{\cdot j} = n_{\cdot j}/n, \tag{10.9}$$

respectivamente. Al sustituir (10.8) y (10.9) en (10.7), se obtiene la estadística

$$\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(N_{ij} - \dfrac{n_{i\cdot}n_{\cdot j}}{n}\right)^2}{\dfrac{n_{i\cdot}n_{\cdot j}}{n}}, \tag{10.10}$$

que para valores grandes de n es, en forma aproximada, una variable aleatoria chi-cuadrada con $(r - 1) \times (c - 1)$ grados de libertad.

Ejemplo 10.5 Una compañía evalúa una propuesta para fusionarse con una corporación. El consejo de directores desea muestrear la opinión de los accionistas para determinar si ésta es independiente del número de acciones que cada uno posee. Una muestra aleatoria de 250 accionistas proporciona la información que se muestra en la tabla 10.4. Con base en esta información, ¿existe alguna razón para dudar de que la opinión con respecto a la propuesta es independiente del número de acciones que posee el accionista? Úsese $\alpha = 0.10$.

La hipótesis nula se establece de la siguiente forma

$$H_0: p_{ij} = p_{i\cdot}p_{\cdot j}, \qquad i = 1, 2, 3; \quad j = 1, 2, 3.$$

En ésta, p_{ij} es la probabilidad de que un accionista seleccionado al azar se encuentre en la categoría (i, j); $p_{i\cdot}$ es la probabilidad marginal de que el número de acciones que posee un accionista seleccionado al azar se encuentre en la categoría i; y $p_{\cdot j}$ es la probabilidad marginal de que un accionista seleccionado al azar tenga una opinión j. Por la expresión (10.10) la frecuencia esperada de la celda (i, j) es el producto del total de i-ésimo renglón por el total de la j-ésima columna dividido por el tamaño de la muestra $n = 250$. Por ejemplo, el número esperado de accionistas que están a favor de la propuesta y que poseen más de 1 000 acciones, es (95)(100)/250 = 38. Al continuar este proceso, se determinan las frecuencias esperadas para cada combinación. En cada celda de la tabla 10.5, la primera línea representa la frecuencia observada, la segunda la frecuencia esperada y la tercera la contribución de cada celda al valor de la estadística, de acuerdo con (10.10).

De esta manera, el valor de la estadística es

$$\chi^2 = \frac{(38 - 30.4)^2}{30.4} + \frac{(29 - 39.52)^2}{39.52} + \cdots + \frac{(4 - 7.6)^2}{7.6} = 10.80.$$

Dado que $r = c = 3$, el número de grados de libertad es 4. Para $\alpha = 0.1$, el valor crítico es $\chi^2_{0.9,4} = 7.78$. De esta forma, el valor que se observa de la estadística de prueba se encuentra dentro de la región crítica, y la hipótesis nula debe rechazarse. De acuerdo con lo anterior, existe una razón para creer que la opinión con respecto a la propuesta y el número de acciones que cada accionista posee, no son independientes.

TABLA 10.4 Datos muestrales para el ejemplo 10.5

Número de acciones	*Opinión*			
	A favor	*En contra*	*Indecisos*	*Totales*
Menos de 200	38	29	9	76
200–1000	30	42	7	79
Más de 1000	32	59	4	95
Totales	100	130	20	250

TABLA 10.5 Frecuencias esperadas y observadas para el ejemplo 10.5

Número de acciones	*A favor*	*En contra*	*Indecisos*	*Totales*
Menos de 200	38 30.40 1.90	29 39.52 2.80	9 6.08 1.40	76 76 6.10
200–1000	30 31.60 0.08	42 41.08 0.02	7 6.32 0.07	79 79 0.17
Más de 1000	32 38 0.95	59 49.40 1.87	4 7.60 1.71	95 95 4.53
Totales	100 100 2.93	130 130 4.69	20 20 3.18	250 250 10.80

Referencias

1. P. G. Hoel, *Introduction to mathematical statistics,* 4th ed., Wiley, New York, 1971.
2. B. W. Lindgren, *Statistical theory*, 3rd ed., Macmillan, New York, 1976.

Ejercicios

10.1. Con base en los registros de una tienda de modas, el 50% de los vestidos adquiridos por ésta para la temporada se venderán a precio de menudeo, el 25% a un 20% menos del precio de menudeo, 15% se venderán después de una reducción en su precio del 40% y los restantes con una disminución en su precio del 60%. Para esta temporada, se adquirieron 300 vestidos y su venta fue en la siguiente forma:

Precio de venta	*20% de*	*40% de*	*60% de*
140	90	30	40.

¿Existe alguna razón para creer que la disminución en ventas fue diferente en esta temporada con respecto a las anteriores? Úsese $\alpha = 0.05$. ¿Cuál es el valor de p?

10.2. En un hospital, el número de nacimientos observados para cada mes de cierto año, fueron los siguientes:

Ene	Feb	Marzo	Abril	Mayo	Jun	Julio	Ago	Sept	Oct	Nov	Dic
95	105	95	105	90	95	105	110	105	100	95	100

Si $\alpha = 0.01$, ¿existe alguna razón para creer que el número de nacimientos no se encuentra distribuido en forma uniforme durante todos los meses del año? ¿Cuál es el valor de p?

10.3. En el ejercicio 10.2, supóngase que el número de nacimientos que se observaron cada mes durante un periodo de 10 años es simplemente igual a diez veces los números observados en el ejercicio 10.2 para un año.

a) ¿Cambiará esto la conclusión del ejercicio 10.2?
b) ¿Qué puede concluirse con respecto al empleo de prueba de bondad de ajuste chi-cuadrada para valores grandes de n?

10.4. Un fabricante asegura que produce sólo el 5% de unidades defectuosas. Un comprador de grandes cantidades de estas unidades selecciona 100 y encuentra diez defectuosas.

a) Mediante el empleo de la prueba de bondad de ajuste chi-cuadrada, determinar si existe una razón para dudar de la afirmación del fabricante. Úsese $\alpha = 0.05$.
b) Compárese la respuesta con la parte *a*, que se obtiene al utilizar el método aproximado que se discutió en el capítulo 9 para probar la hipótesis nula de que la verdadera proporción de artículos defectuosos es 0.05.
c) ¿Existe alguna relación entre los valores de las estadísticas de prueba obtenidos en las partes *a* y *b*? ¿Existe alguna condición para esta relación?

10.5. Una organización de seguridad vial desea determinar si el número de accidentes fatales se encuentra distribuido de igual forma para el color de los automóviles involucrados en los accidentes. La organización obtuvo una muestra aleatoria de 600 accidentes automovilísticos en los cuales ocurrió por lo menos una muerte y anotó el color del automóvil. Se obtuvo la siguiente información:

Rojo	*Café*	*Amarillo*	*Blanco*	*Gris*	*Azul*
75	125	70	80	135	115

¿Existe alguna razón para creer que las proporciones de color no son idénticas? Úsese $\alpha = 0.01$.

10.6. Durante un periodo de 30 años se llevó a cabo un estudio médico para determinar, entre otras cosas, si los hábitos de fumador pueden influenciar en el desarrollo de la enfermedad cardiaca. Durante este periodo, 160 hombres desarrollaron alguna enfermedad cardiaca. Estos hombres fueron clasificados como fumadores agudos (más de dos cajetillas de cigarros al día), fumadores moderados (una a dos cajetillas al día), fumadores ocasionales (menos de una cajetilla al día) o no fumadores. El número de hombres en cada categoría que desarrolló alguna enfermedad cardiaca es el siguiente:

Fumador agudo	*Fumador moderado*	*Fumador ocasional*	*No fumador*
58	54	36	12

a) Si se supone que al comienzo del estudio había una cantidad igual de hombres en cada una de las cuatro categorías, ¿existe alguna razón a un nivel de $\alpha = 0.01$ para creer que las proporciones en estas categorías no son las mismas?
b) ¿Cómo se podría prevenir al investigador médico del uso de la prueba de bondad de ajuste chi-cuadrada en esta situación?

10.7. En un proceso de producción se toma una muestra aleatoria diaria de 100 artículos y se inspecciona para encontrar artículos defectuosos. Para una semana dada y para los cinco días de operación, se observó el siguiente número de unidades defectuosas:

Lunes	*Martes*	*Miércoles*	*Jueves*	*Viernes*
12	7	6	5	10

Si el porcentaje total de artículos defectuosos es del 8%, ¿puede concluirse que a un nivel de $\alpha = 0.05$ existe una diferencia discernible en el porcentaje diario de artículos defectuosos?

10.8. Con referencia a los datos del ejercicio 1.1, empleando la prueba de bondad de ajuste chi-cuadrada, ¿puede concluirse que los lapsos de tiempo no se encuentran exponencialmente distribuidos con $\theta = 3.2$ minutos? Úsese $\alpha = 0.01$.

10.9. Considere los datos del ejercicio 1.7.

a) Para $\alpha = 0.05$, empléese la prueba de bondad de ajuste chi-cuadrada para probar la hipótesis nula de que la distribución del número de anotaciones de seis puntos por equipo y por juego en la NFL, es una distribución de Poisson con parámetro $\lambda = 2.7$.

b) Supóngase que se estima el valor de λ a partir de los datos. ¿Cómo podría este cambio efectuar la respuesta a la parte *a*?

10.10. Úsese la estadística de Kolmogorov-Smirnov en los datos del ejercicio 1.1 y compare el resultado con el que se obtiene en el ejercicio 10.8.

10.11. Úsese lá estadística de Kolmogorov-Smirnov para probar la hipótesis nula de que los datos del ejercicio 1.2 se encuentran normalmente distribuidos con media 50 y desviación estándar 10. Úsese $\alpha = 0.05$.

10.12. Como se notó con anterioridad, una limitación de la estadística de Kolmogorov-Smirnov es que debe especificarse el modelo propuesto bajo H_0. A pesar de que no se encuentra disponible ningún método cuando algunos de los parámetros no se especifica, Lilliefors* obtuvo los límites de rechazo a través de un estudio de simulación para el problema específico de probar la normalidad. Si la media y la desviación estándar muestral se emplean como parámetros de la distribución normal bajo la hipótesis nula, la estadística D_n tiene una distribución cuyos cuantiles también obtuvo Lilliefors. De manera específica, para $\alpha = 0.05$ los valores del 95avo, percentil de la distribución de esta estadística bajo H_0 fueron los siguientes:

n	10	12	14	15	16	18	20	25	>25
95avo. percentil	0.258	0.242	0.227	0.220	0.213	0.200	0.190	0.173	$0.886/\sqrt{n}$

Empléese la modificación de Lilliefors a la estadística de Kolmogorov-Smirnov para probar la normalidad de los datos del ejercicio 1.2. Compárese el resultado con el del ejercicio 10.11.

10.13. Úsese el procedimiento de la prueba de bondad de ajuste chi-cuadrada para probar la hipótesis nula de que los datos del ejercicio 1.2 se encuentran distribuidos, normalmente, a un nivel de $\alpha = 0.01$.

10.14. Se toma una muestra aleatoria de 25 hombres casados y se les pregunta la edad que tenían cuando se casaron. Se obtienen los siguientes datos: 24, 19, 20, 22, 50, 23, 23,

On the Kolmogorov–Smirnov test for normality with mean and variance unknown*. J. Amer. Statistical Assoc. **64 (1967), 399–402. 1967.

21, 25, 27, 45, 27, 26, 26, 35, 29, 28, 30, 31, 32, 31, 33, 34, 38, 41. Úsese la estadística de Kolmogorov-Smirnov para probar la hipótesis nula de que la distribución de las edades de los hombres cuando contrajeron sus primeras nupcias es una distribución gama con $\theta = 2$ y $\alpha = 16$. Úsese $\alpha = 0.05$. (Sugerencia: Para calcular las probabilidades gama, véase una tabla de la función gama incompleta determinada por (5.55).)

10.15. En el ejemplo 4.10, úsese la prueba de bondad de ajuste chi-cuadrada para demostrar que la hipótesis nula de una distribución binomial negativa para el número de anotaciones de seis puntos, no puede ser rechazada a un nivel $\alpha = 0.05$.

10.16. Con la prueba de bondad de ajuste chi-cuadrada determínese si la hipótesis nula de los datos del accidente del ejercicio 8.14 sigue una distribución binomial negativa, que se puede remitir al nivel $\alpha = 0.05$

10.17. Los totales de los renglones y columnas de una tabla de contingencia de dos características son los siguientes:

				10
				12
				15
8	14	10	5	37

Bajo la hipótesis nula de independencia, determinar la tabla de frecuencias esperadas.

10.18. Un proceso de producción emplea cinco máquinas en sus tres operaciones de desplazamiento. Se clasificó una muestra aleatoria de 164 fallas de acuerdo con la máquina y la operación de desplazamiento en la que ocurrió la falla, y los resultados se muestran en la tabla 10.6. Con base en esta información, ¿existe alguna razón para dudar acerca de la independencia entre la operación de desplazamiento y la falla de la máquina? Úsese $\alpha = 0.01$.

TABLA 10.6 Fallas por máquina y desplazamiento

	Máquinas				
Desplazamiento	*A*	*B*	*C*	*D*	*E*
1	10	12	8	14	8
2	15	8	13	8	11
3	12	9	14	12	10

10.19. Se condujo una encuesta aleatoria entre los ciudadanos en edad de votar para determinar si existía alguna relación entre la afiliación partidista y la opinión con respecto al control de armas. Se obtuvo la información proporcionada en la tabla 10.7. Para $\alpha = 0.01$, ¿existe alguna razón para creer que existe una dependencia entre la opinión y la afiliación partidista?

TABLA 10.7 Filiación partidaria y opiniones sobre el control de armas

	A favor	*En contra*	*Sin decisión*
Demócratas	110	64	26
Republicanos	90	116	14
Independientes	55	35	10

10.20. En una muestra aleatoria de recién egresados de la preparatoria se registraron dos características (la calificación promedio y el número de respuestas correctas para la prueba SAT). Esta información se clasificó como se muestra en la tabla 10.8

TABLA 10.8 Calificaciones promedio y número de respuestas correctas para la prueba SAT

	Número de respuestas correctas para la prueba SAT		
GPA	900–1100	1100–1300	1300–1500
>3.5	50	65	38
3.0–3.5	78	72	42
2.5–3.0	97	80	25
2.0–2.5	105	25	18

a) ¿Existe una dependencia entre el número de respuestas correctas en la prueba SAT y el promedio de clasificaciones. discernible estadísticamente a un nivel $\alpha = 0.01$?

b) ¿Se tiene alguna reserva con respecto a esta clasificación? ¿Se puede pensar en otras características que deban considerarse?

10.21. En un estudio reciente que involucró una muestra aleatoria de 300 accidentes automovilísticos, se clasificó la información de acuerdo con el tamaño del automóvil.

	Pequeño	*Mediano*	*Grande*
Por lo menos un muerto	42	35	20
Ningún muerto	78	65	60

Con estos datos, ¿depende la frecuencia de accidentes del tamaño del automóvil? Úsese $\alpha = 0.05$.

10.22. Se llevó a cabo una encuesta con respecto a la preferencia del consumidor para determinar si existía alguna predilección para tres marcas competitivas (A, B y C) dependiendo de la región geográfica en la que habita el consumidor. Con base en una muestra aleatoria de consumidores, se obtuvo la siguiente información para tres distintas regiones.

	Región 1	*Región 2*	*Región 3*
Marca A	40	52	25
Marca B	52	70	35
Marca C	68	78	60

Con base en esta información, ¿la preferencia por una determinada marca depende de la región geográfica a un nivel $\alpha = 0.05$?

CAPÍTULO ONCE

Métodos para el control de calidad y muestreo para aceptación

11.1 Introducción

En los últimos años ha aumentado el interés que se tiene, por parte de los productores así como de los consumidores, en la calidad de los productos manufacturados. Un fabricante que desea mantener cierto nivel de calidad en su producto terminado debe implantar un procedimiento para detectar cualquier desviación seria del estándar de calidad deseado. En el logro de este fin, las tablas estadísticas de control de calidad y el muestreo periódico han demostrado ser medios muy efectivos para controlar la calidad de los productos manufacturados.

Por otro lado, el consumidor desea asegurarse de que el producto que adquiere reúne ciertos estándares de calidad. Lo anterior es especialmente cierto si el consumidor, como muchas veces ocurre en la práctica, compra lotes muy grandes de cierto producto. En estos casos es necesario establecer un procedimiento para inspeccionar una muestra relativamente pequeña del producto proveniente del lote para decidir si reúne los estándares de calidad deseados. Un procedimiento de esta naturaleza incluye la noción del muestreo para aceptación.

En este capítulo se analizarán los principios básicos y métodos de las tablas de control estadístico y los procedimientos del muestreo para aceptación. El lector debe considerar el material de este capítulo sólo como introducción al control estadístico de calidad y a los procedimientos del muestreo para aceptación, pero éste debe ser útil como antecedente para un estudio posterior. Con este propósito se sugieren las referencias [2] y [3].

11.2 Tablas de control estadístico

Una tabla de control estadístico es un procedimiento inferencial basado en un muestreo repetitivo para estudiar un proceso. De acuerdo con su creador, W.A.

Shewhart, una tabla de control se emplea para definir un estándar de calidad para un proceso de fabricación y para determinar si éste se mantiene por el proceso.

En el desarrollo de tablas de control, el factor clave es la variabilidad en la calidad del producto terminado. Para cualquier proceso, es inherente cierta cantidad de variabilidad en la calidad, sin importar cuántos esfuerzos se encaminen para lograr su control. Este tipo de variabilidad es una función de factores aleatorios que, de manera común, se encuentran más allá del control. Esta variación aleatoria generalmente es aceptable y no compromete en modo alguno el estándar de calidad deseado. La variabilidad también se puede deber a causas no aleatorias o fijas; éstas pueden tomar la forma de un mal funcionamiento en una máquina, indiferencia del trabajador, variabilidad en la calidad de las materias primas y otras. De esta forma, una tabla de control estadístico es el procedimiento inferencial con el cual se decide si una desviación observada de la norma deseada se debe sólo al azar o a alguna causa fija. Si la decisión es que la variación es aleatoria, entonces se dice que el proceso de interés se encuentra bajo control. De otro modo, se juzga como fuera de control y en este caso lo que se hace, en forma general, es detener el proceso y llevar a cabo todos los esfuerzos necesarios para detectar la causa del problema.

Dado que la inferencia se basa en la probabilidad, es posible que un proceso se juzgue fuera de control cuando, de hecho, se encuentra bajo control o viceversa. Las consecuencias de estos errores pueden ser severas; por ejemplo si se declara a un proceso como fuera de control, cuando en realidad está bajo control, se tratará de determinar una causa inexistente. Por otro lado, si el proceso en realidad está fuera de control y se permite que éste continúe, el estándar de calidad deseado no se alcanzará. Debe notarse que estos errores son facsímiles de los errores de tipo I y II analizados en el capítulo 9.

Usualmente, la determinación de una tabla de control depende de la toma periódica de muestras aleatorias de tamaño n del proceso de interés, con lo que se obtiene, para cada una de éstas, un valor de alguna estadística de importancia como la media o la varianza muestral. Por lo tanto, la tabla de control es una gráfica de los valores de la estadística observada, contra el número de la muestra o contra el periodo durante el cual se obtuvo ésta. La tabla contiene límites de control superior e inferior, los cuales constituyen los criterios de decisión para el proceso, es decir, el proceso será juzgado como bajo control mientras los valores de la estadística se encuentren dentro de estos límites. Si un valor de la estadística se encuentra fuera de los límites de control, se considerará al proceso como fuera de control. También se encuentra una línea central que define la norma prescrita para el proceso.

El usuario decide cuáles deben ser los valores de los límites de control, cuántas veces es necesario muestrear, cuál debe ser el tamaño de la muestra que se toma y qué acción realizar una vez que se juzga al proceso como fuera de control. Sin embargo, existen algunos principios generales que el usuario puede seguir. Shewhart argumentaba que podía alcanzarse un balance apropiado entre el costo del muestreo y la exactitud del estimador, si las muestras tienen un tamaño de cuatro o cinco observaciones cada vez. También los límites de control "tres-sigma" han demostrado ser muy satisfactorios y son los que se emplean en Estados Unidos, así como en muchos otros países.

Considérense las tablas de control para la media y la desviación estándar. La primera se conoce como tabla $\overline{X}$ y la segunda como tabla S. Debe notarse que, de ma-

nera tradicional, se emplea el rango R para determinar tablas para la variabilidad de un proceso debido a su cálculo fácil. Pero es mejor la tabla S, la cual no ofrece ningún problema de cálculo con los paquetes para computadora disponibles en la actualidad. Para la determinación de las tablas $\overline{X}$ y S se supondrá que se muestrea una distribución normal; en un caso, se dará por hecho que se conoce el valor de la media o el de la varianza y, para el otro, que ambos valores son desconocidos.

11.2.1 Tablas $\overline{X}$ (media conocida de la población)

Se puede construir una tabla de control con base en la media muestral cuando la medición de interés se encuentra normalmente distribuida con media μ y desviación estándar o σ conocidas. El conocimiento que se tiene sobre μ y σ se puede deber a la naturaleza particular del proceso de interés, el cual puede proporcionar la suficiente información con respecto a la media y a la desviación estándar. Para este caso, una tabla $\overline{X}$ proporciona el procedimiento inferencial por medio del cual se puede decidir si la media del proceso es la que se afirma.

Sea $X_1, X_2, \ldots, X_n$ una muesta aleatoria de tamaño n del proceso de interés. Dado que por hipótesis $X_i \sim N(\mu, \sigma)$, la media muestral es $\overline{X} \sim N(\mu, \sigma/\sqrt{n})$, la probabilidad de que $|\overline{X} - \mu|$ sea menor que $3\sigma/\sqrt{n}$, es

$$P(|\overline{X} - \mu| < 3\sigma/\sqrt{n}) = 0.9974.$$

De esta forma, los límites de control tres-sigma son $\mu \pm 3\sigma/\sqrt{n}$, es decir, cuando se toma una muestra de tamaño n se calcula y se grafica un valor de la media muestral. Si éste se encuentra dentro de los límites de control $\mu \pm 3\sigma/\sqrt{n}$, se supone que el proceso se encuentra bajo control; de otra forma, está fuera de control. Por lo tanto, cada vez que se toma una muestra se está probando la hipótesis nula de que la media del proceso es igual a μ contra la alternativa de que ha ocurrido un corrimiento en la media del proceso. El rechazo de la hipótesis nula implica que el proceso se encuentra fuera de control.

Ejemplo 11.1 En un proceso de llenado se tiene una máquina que vacía una cantidad promedio de 500 g en cada recipiente, con una desviación estándar de 2 g. Se toman 10 muestras diarias, cada una de cinco recipientes, y se mide el peso de cada recipiente. Los pesos promedio para las 10 muestras en una semana dada son los siguientes:

Número de muestra	1	2	3	4	5
Promedio de la muestra	498.37	499.49	501.25	498.63	502.97

Número de muestra	6	7	8	9	10
Promedio de la muestra	500.56	499.23	498.76	501.05	500.27

Para los límites de control 3σ, ¿se encontró el proceso bajo control durante esta semana? Con estos límites, ¿cuál es la probabilidad de no detectar un corrimiento de 500 a 503 g en la media?

Dado que $n = 5$, $\mu = 500$, y $\sigma = 2$, los límites de control 3σ son $500 \pm 3(2/\sqrt{5}) = 500 \pm 2.6833$ o (497.3167, 502.5833). En la figura 11.1 se muestra la tabla de control para las medias muestrales. Nótese que la quinta media muestral se encuentra por encima del límite superior de control; de esta forma, durante este tiempo el proceso se juzgó como fuera de control en relación con el promedio. La probabilidad de observar un valor de $\bar{X}$ fuera de los límites de control, si el proceso se encuentra realmente bajo control, es

$$P(|\bar{X} - 500| > 2.6833) = 0.0026.$$

La probabilidad de no detectar un corrimiento de 500 a 503 gramos en la media es

$$\begin{aligned} P(497.3167 < \bar{X} < 502.6833 \mid \mu = 503) &= P\left(\frac{497.3167 - 503}{2/\sqrt{5}} < Z < \frac{502.6833 - 503}{2/\sqrt{5}}\right) \\ &= P(-6.35 < Z < -0.35) \\ &= 0.3632. \end{aligned}$$

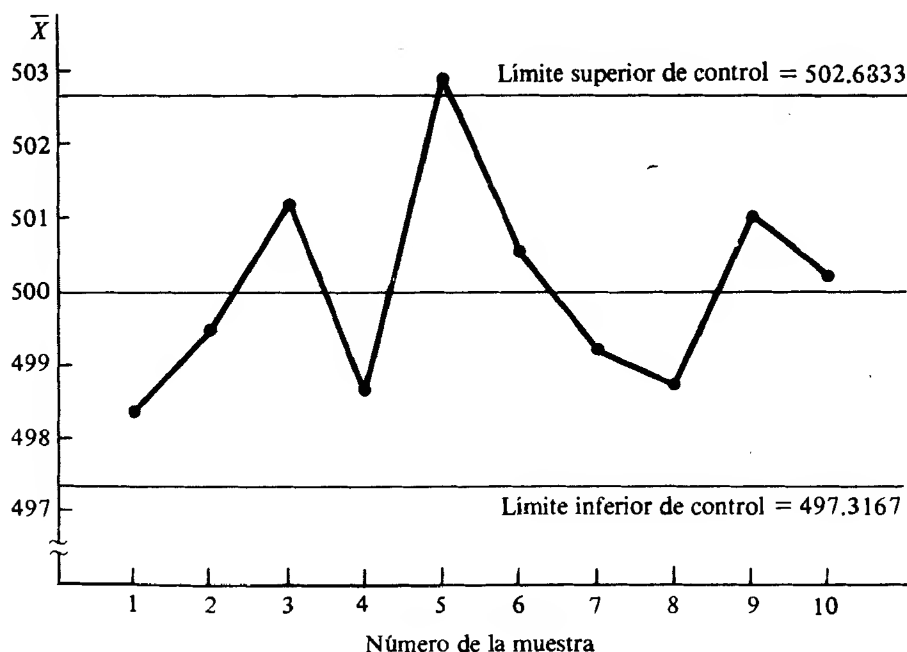


FIGURA 11.1 Tabla $\bar{X}$ para los datos del ejemplo 11.1

11.2.2 Tablas S (desviación estándar conocida de la población)

En muchas ocasiones la variabilidad de un proceso es, por lo menos, tan importante como la media de éste; por ejemplo, en la fabricación de instrumentos de precisión, mantener la variación en las mediciones a un nivel aceptable es, probablemente, tan importante como el promedio.

Se considerarán las tablas de control para la variabilidad de un proceso mediante el empleo de la desviación estándar de la muestra

$$S = \left[\sum(X_i - \overline{X})^2/(n-1)\right]^{1/2}.$$

Los límites de control 3σ son $E(S) \pm 3\,d.e.(S)$. Para obtener $E(S)$ y $Var(S)$, recuérdese de la sección 7.5 que la variable aleatoria

$$Y = \frac{(n-1)S^2}{\sigma^2}$$

tiene una distribución chi-cuadrada con $n-1$ grados de libertad, en donde la función de densidad de probabilidad de Y está dada por (7.16). Dado que

$$S^2 = \frac{\sigma^2 Y}{n-1},$$

entonces

$$S = \frac{\sigma Y^{1/2}}{(n-1)^{1/2}},$$

y

$$E(S) = \frac{\sigma}{(n-1)^{1/2}} E(Y^{1/2}).$$

Pero

$$E(Y^{1/2}) = c\int_0^\infty y^{1/2}\, y^{(n-3)/2} \exp(-y/2)dy, \tag{11.1}$$

en donde

$$c = \frac{1}{\Gamma[(n-1)/2]\, 2^{(n-1)/2}}.$$

En (11.1) sea $u = y/2$; entonces $dy = 2du$ y

$$E(Y^{1/2}) = 2^{n/2}\, c\int_0^\infty u^{(n-2)/2} \exp(-u)du = 2^{n/2}c\Gamma(n/2).$$

Entonces

$$E(S) = \frac{\sigma}{(n-1)^{1/2}}\, 2^{n/2}\, c\Gamma(n/2)$$

$$= \sigma\frac{2^{1/2}\Gamma(n/2)}{(n-1)^{1/2}\Gamma[(n-1)/2]}. \tag{11.2}$$

Es preferible utilizar una notación para el control de calidad y escribir

$$E(S) = c_4\sigma,$$

en donde

$$c_4 = \frac{2^{1/2}\Gamma(n/2)}{(n-1)^{1/2}\Gamma[(n-1)/2]}. \tag{11.3}$$

Para la varianza de S, por definición

$$Var(S) = E(S^2) - E^2(S).$$

Pero en la sección 7.5 se demostró que $E(S^2) = \sigma^2$, en consecuencia

$$Var(S) = \sigma^2 - c_4^2\sigma^2 = \sigma^2(1 - c_4^2),$$

o en la notación preferible,

$$Var(S) = c_5^2\sigma^2.$$

Por lo tanto, $d.e.(S) = c_5\sigma$, y los límites de control 3σ son

$$c_4\sigma \pm 3c_5\sigma, \tag{11.4}$$

en donde c_4 está dada por (11.3) y $c_5 = (1 - c_4^2)^{1/2}$. Nótese que, dado que se supone que el valor de σ se conoce, los límites de control sólo son funciones del tamaño de cada muestra. En la tabla 11.1 se determinan los valores de c_4 y c_5 para distintos valores usuales del tamaño n de las muestras.

Como ilustración, si $\sigma = 2$, los límites de control 3σ para la desviación estándar muestral, con base en $n = 5$, son $(0.94)(2) \pm (3)(0.3412)(2)$ o $(0, 3.9272)$. Para este ejemplo, en la tabla S el límite inferior de control es cero, la línea central se encuentra en 1.88 y el límite superior de control es 3.9272. Para $n = 5$ y $\sigma = 2$, la variabilidad del proceso se considera bajo control, siempre que el valor de la desviación estándar muestral se encuentre dentro de los límites de control ya establecidos.

11.2.3 Tablas $\bar{X}$ y S (media y varianza desconocidas de la población)

Se considerarán las tablas de control para aquellos casos en los que la distribución de la población es normal, pero no se conocen los valores de la media y la desviación estándar. Para esta situación, los límites de control se basan en los valores estimados para μ y σ.

Dado que no se conoce el valor promedio del proceso, tampoco se conoce la línea central de la tabla de control. Si la línea central es un valor estimado basado en un gran número de muestras, los límites de control que se obtienen de esta manera de-

TABLA 11.1 Valores de c_4 y c_5 para tamaños n normales de la muestra

n	4	5	6	7	8	9	10
c_4	0.9213	0.9400	0.9515	0.9594	0.9650	0.9693	0.9727
c_5	0.3889	0.3412	0.3076	0.2820	0.2622	0.2459	0.2321

ben considerarse sólo como *límites tentativos*, ya que quizá se necesite una modificación antes de que se puedan utilizar para medir la calidad de un producto en futuras operaciones de producción. Lo anterior significa que los límites de control tentativos son apropiados para determinar si las operaciones pasadas de un proceso de producción estuvieron bajo control. Para extenderlos a la producción futura, el procedimiento usual es eliminar todos aquellos puntos que se encuentren fuera de los límites tentativos de control y recalcular el valor de éstos con base en el resto de la información muestral. Se continúa este proceso hasta que todos los puntos se encuentren dentro de los límites de control, tanto para la tabla $\overline{X}$ como para S. La razón para este procedimiento es que los límites de control para la futura producción deben ser funciones de las observaciones que se recabaron mientras el proceso de producción estaba bajo control.

De acuerdo con Shewhart, los límites tentativos de control deben estar basados, por lo menos, en 20 muestras, cada una con cuatro o cinco observaciones. Shewhart denominó a estas muestras *subgrupos racionales*. Éstos deben seleccionarse de manera tal que cada subgrupo sea prácticamente homogéneo y proporcione la máxima oportunidad de variación de un subgrupo a otro. Para un proceso de producción esto implica que las observaciones para un subgrupo deben tomarse en un momento que sea diferente al de otro subgrupo. Se emplea un tamaño relativamente pequeño de la muestra de cuatro o cinco observaciones, no sólo para mantener el balance entre el costo del muestreo y la exactitud del estimado, sino también para dar una mínima oportunidad de variación dentro de cada subgrupo.

Sea m el número de muestras y supóngase que $n_i = n$ para toda $i = 1, 2, \ldots, m$. Además, sean $\overline{X}_i$ y S_i la media y desviación muestral de la i-ésima muestra. Para todas las m muestras, defínanse las estadísticas.

$$\overline{\overline{X}} = \frac{1}{m}\sum_{i=1}^{m}\overline{X}_i \tag{11.5}$$

y

$$\overline{S} = \frac{1}{m}\sum_{i=1}^{m}S_i. \tag{11.6}$$

Es evidente que $E(\overline{\overline{X}}) = \mu$; de esta forma, el promedio de todas las m muestras en un estimador no sesgado de μ. De manera similar,

$$E(\overline{S}) = \frac{1}{m}\sum E(S_i) = \frac{1}{m}(mc_4\sigma) = c_4\sigma,$$

lo cual sugiere que un estimador de σ es $\overline{S}/c_4$. Los límites tentativos 3σ para la media muestral cuando no se conocen los valores de μ y σ son

$$\overline{\overline{X}} \pm 3\frac{\overline{S}}{c_4\sqrt{n}}, \tag{11.7}$$

y los correspondientes a la desviación estándar de muestra son

$$\overline{S} \pm 3\frac{c_5\overline{S}}{c_4}, \tag{11.8}$$

en donde los valores de c_4 y c_5 son los ya definidos.

Ejemplo 11.2 Los datos en la tabla 11.2 son 20 muestras, cada una con cinco observaciones tomadas en intervalos de dos horas, de la resistencia a la tensión en libras de un hilo. Para cada muestra se proporcionan los valores de la media y la desviación estándar. Constrúyanse las tablas de control $\overline{X}$ y S con base en estos datos.

Al promediar las 20 medias muestrales se obtiene $\bar{\bar{x}} = 47.12$, y si se promedian las desviaciones estándar muestrales, se tiene $\bar{s} = 2.326$. Para $n = 5$, $c_4 = 0.94$ y $c_5 = 0.3412$. Entonces, por (11.7) y (11.8), los límites tentativos de control 3σ para las medias muestrales son

$$47.12 \pm \frac{(3)(2.326)}{(0.94)\sqrt{5}} = (43.80, 50.44),$$

y los límites para las desviaciones estándar muestrales son

$$2.326 \pm \frac{(3)(0.3412)(2.326)}{0.94} = (0, 4.8589).$$

En la figura 11.2 se proporcionan las tablas de control. Nótese que la variabilidad del proceso parece estar bajo control, pero la media muestral para la vigésima muestra se encuentra fuera de los límites tentativos. Debido a lo anterior, se obtienen nuevos valores para los límites después de omitir esta muestra. Éstos son

$$47.31 \pm \frac{(3)(2.368)}{(0.94)\sqrt{5}} = (43.93, 50.69)$$

TABLA 11.2 Datos de la muestra de la resistencia a la tensión de un hilo en libras

Número de la muestra			*Valores de la muestra*			$\overline{X}$	S
1	44	46	48	52	49	47.8	3.03
2	44	47	49	46	44	46.0	2.12
3	47	49	47	43	44	46.0	2.45
4	45	47	51	46	48	47.4	2.30
5	44	41	50	46	50	46.2	3.90
6	49	46	45	46	49	47.0	1.87
7	47	48	50	46	47	47.6	1.52
8	49	46	51	48	46	48.0	2.12
9	47	42	48	44	46	45.4	2.41
10	46	48	45	51	50	48.0	2.55
11	45	47	51	48	46	47.4	2.30
12	52	51	48	48	45	48.8	2.77
13	45	45	47	49	44	46.0	2.00
14	46	47	43	48	45	45.8	1.92
15	48	49	52	46	51	49.2	2.39
16	44	46	45	47	52	46.8	3.11
17	48	50	47	46	49	48.0	1.58
18	48	52	51	47	46	48.8	2.59
19	47	51	50	46	49	48.6	2.07
20	44	43	42	43	46	43.6	1.52

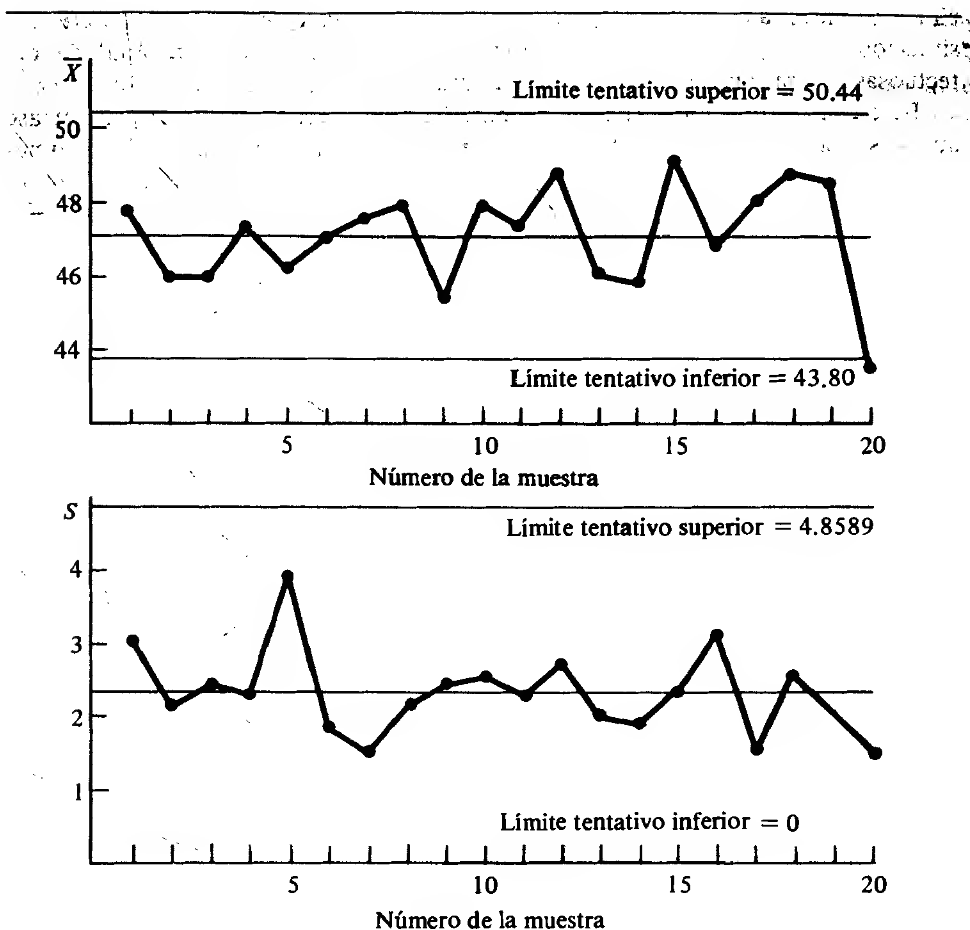


FIGURA 11.2 Tablas $\overline{X}$ y S para los datos del ejemplo 11.2

para $\overline{X}$ y los límites

$$2.368 \pm \frac{(3)(0.3412)(2.368)}{0.94} = (0 \quad 4.9466)$$

para S. Se observa que todos los puntos se encuentran dentro de los nuevos límites tentativos, tanto en la tabla $\overline{X}$, como en la S.

La construcción de las tablas $\overline{X}$ y S se basa en la distribución normal. La tabla $\overline{X}$ es, relativamente, insensible a la hipótesis de normalidad debido al teorema del límite central. Sin embargo, la tabla S es mucho más sensible a la hipótesis de normalidad.

Vale la pena mencionar la existencia de la tabla p. La tabla p puede construirse cuando se supone que el muestreo se lleva a cabo sobre una distribución binomial con parámetro de proporción p. Los límites de control se obtienen para las propor-

ciones de muestra de unidades que caen en una de dos categorías posibles. Para esta situación, lo que generalmente es de interés, es vigilar la proporción de unidades defectuosas que produce un proceso de manufactura.

Para construir los límites de control para las proporciones muestrales, supóngase que no se conoce el valor de p. Sea m el número de muestras disponible, y X_i el número de unidades defectuosas en la i-ésima muestra de tamaño n. Entonces X_i/n es un estimador de p basado en la i-ésima muestra, y $\overline{P} = (1/mn) \sum_{i=1}^{m} X_i$ es un estimador de p basado en todas las m muestras. De acuerdo con lo anterior, los límites tentativos 3σ para las proporciones muestrales X_i/n son

$$\overline{P} \pm 3\sqrt{\frac{\overline{P}(1 - \overline{P})}{n}}. \tag{11.9}$$

11.3 Procedimientos del muestreo para aceptación

Un consumidor puede escoger uno de los tres caminos siguientes para verificar la calidad de los artículos de un embarque que ha recibido: inspeccionar todos los artículos en el lote; inspeccionarlos en una muestra aleatoria proveniente del lote, o aceptar el lote sin llevar a cabo ninguna inspección. La primera opción tiene generalmente un precio prohibitivo y la última es poco probable que sea aceptada por un consumidor serio, con respecto a la calidad de los artículos que adquiere. Por lo tanto, la opción que tiene un balance adecuado entre el costo de la inspección y el que implica aceptar un lote y usar artículos defectuosos, es la de inspeccionar los artículos en una muestra aleatoria proveniente del lote que se acaba de adquirir. Con base en el proceso de inspección, la decisión usual es aceptar el lote, rechazarlo o tomar otra muestra aleatoria. Si la decisión de aceptar o rechazar se toma con base en los valores medidos de los artículos, con respecto a una medición física continua, entonces se dice que la inspección se lleva a cabo *por variables*. Si los artículos que se inspeccionan se clasifican como defectuosos o no defectuosos, y el lote se acepta o rechaza con base en el número de artículos defectuosos en la muestra, se dice que la inspección se lleva a cabo *por características*.

En esta sección se considerarán los fundamentos para desarrollar planes sencillos de muestreo con base en características para decidir si se acepta o se rechaza un lote. Posteriormente se examinará en forma breve el muestreo para aceptación por variables. Sea N el tamaño del lote. Entonces un plan básico de muestreo para aceptación es seleccionar n artículos del lote de tamaño N y aceptar el lote si el número de artículos defectuosos en la muestra es menor o igual a un número de aceptación c, previamente estipulado. De otra forma, el lote se rechaza. Por ejemplo, un plan de muestreo puede definirse de la siguiente forma $N = 10\,000$, $n = 100$, y $c = 1$. Lo anterior significa que se seleccionarán, en forma aleatoria, 100 artículos de los 10 000 que contiene el lote, y si se encuentra cuando mucho un artículo defectuoso, se aceptará el lote de $N = 10\,000$ artículos. Si hay más de un artículo defectuoso, el lote será rechazado. El consumidor puede escoger entre regresar el lote rechazado al fabricante o someterlo a una inspección del 100%. El primero constituye lo que se conoce como un procedimiento de *inspección no verificable*, y el segundo como proceso de *inspección verificable*.

Supóngase que la información disponible para el consumidor con respecto a la calidad de los artículos en el lote, es la proporción promedio de artículos defectuosos que produce el proceso de manufactura que los fabrica. Un criterio muy importante en un plan de muestreo es la probabilidad de aceptar el lote $P(A)$, dada una proporción de artículos defectuosos p. Bajo las hipótesis adecuadas y para algún valor de p y de c, la probabilidad de que el lote sea aceptado con base en una muestra de tamaño n, es la probabilidad binomial acumulativa

$$P(A) \equiv P(X \leq c) = \sum_{x=0}^{c} \binom{n}{x} p^x (1 - p)^{n-x}, \tag{11.10}$$

en donde la variable aleatoria X representa el número de artículos defectuosos encontrados en la muestra. Si np tiene un tamaño moderado, la probabilidad binomial dada por (11.10) se puede aproximar en forma adecuada por la probabilidad acumulativa de Poisson

$$P(A) = \sum_{x=0}^{c} \frac{\lambda^x}{x!} \exp(-\lambda), \tag{11.11}$$

en donde $\lambda = np$.

Una gráfica de la probabilidad de aceptación contra p, es la curva de operación característica (CO). Como ilustración se analizará el plan de muestreo $n = 100$ y $c = 2$. Mediante el empleo de la aproximación de Poisson dada por (11.11) se obtiene la probabilidad de aceptar para valores de p en un intervalo de 0.01 a 0.09. Las probabilidades de aceptación se dan en la tabla 11.3 y están graficadas contra p en la figura 11.3.

La naturaleza de una curva *CO* es afectada por el tamaño n de la muestra y por el número de aceptación c. Como ilustración, considérense los planes de muestreo $n = 50$, $c = 1$; $n = 100$, $c = 2$; y $n = 200$, $c = 4$. En la figura 11.4 se muestran las curvas *CO* para estos planes. Nótese que aunque el cociente c a n es constante, las curvas *CO* son algo diferentes. De hecho, las curvas son más sensibles al tamaño de la muestra. Conforme n aumenta, la pendiente de la curva se torna más pronunciada. De esta forma, para tamaños grandes de la muestra, la probabilidad de aceptación disminuye muy rápidamente conforme el valor de p aumenta. Si el valor de n es fijo, un aumento en el número de aceptación c tenderá a desplazar a la curva hacia la derecha. Esto implica que para una p dada, la probabilidad de aceptación es alta conforme c aumenta. En consecuencia, puede pensarse que entre más cercano a cero se encuentre el valor de c, mejor es el plan de muestreo. Pero la figura 11.4 indica que los planes con valores grandes de c son mejores siempre que el tamaño de la muestra sea, apreciablemente, grande.

TABLA 11.3 Probabilidades de aceptación para el plan de muestreo $n = 100$, $c = 2$

p	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
$P(A)$	0.9197	0.6767	0.4232	0.2381	0.1247	0.0620	0.0296	0.0138	0.0062

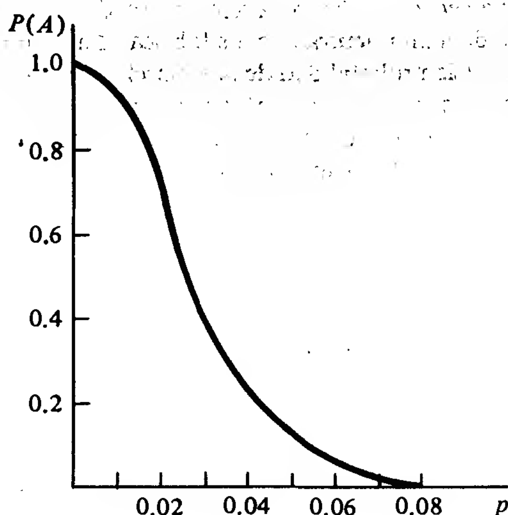


FIGURA 11.3 Curva característica de operación para el plan de muestreo $n = 100$, $c = 2$

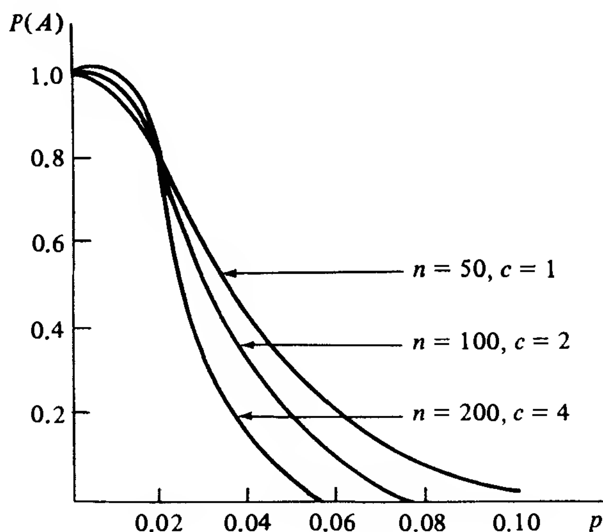


FIGURA 11.4 Curvas características de operación para los tres planes de muestreo

El desarrollo de buenos planes de muestreo incluye tanto al productor como al comprador del lote. De manera normal el productor es el vendedor y el consumidor el comprador. Un productor ciertamente desearía que el consumidor rechazara un porcentaje muy pequeño de los lotes vendidos y que son, en general, buenos; el consumidor desearía aceptar un porcentaje muy pequeño de los lotes que son malos. De esta forma los dos experimentan cierto riesgo. Supóngase que ambos están de acuerdo en que un lote es aceptable si la proporción de artículos defectuosos es $p \leq p_1$, y no aceptable si $p \geq p_2$. Se dan las siguientes definiciones que implican riesgos.

Definición 11.1 El riesgo del productor α es la probabilidad de que el consumidor rechace un lote cuya proporción de artículos defectuosos no es mayor que p_1.

Definición 11.2 El riesgo del consumidor β es la probabilidad de aceptar un lote cuya proporción de artículos defectuosos es mayor o igual a p_2.

Con base en estas definiciones, el riesgo del productor es la probabilidad del error de tipo I, dado que éste representa la probabilidad de rechazar un lote aceptable. De manera similar, el riesgo del consumidor es la probabilidad del error de tipo II, ya que éste representa la probabilidad de equivocarse al no rechazar un lote inaceptable. En otras palabras, la situación anterior es análoga a probar la hipótesis nula H_0: $p = p_1$ contra la alternativa H_1: $p = p_2$.

Los riesgos del productor y del consumidor pueden representarse por dos puntos sobre una curva característica de operación, como se ilustra en la figura 11.5. En

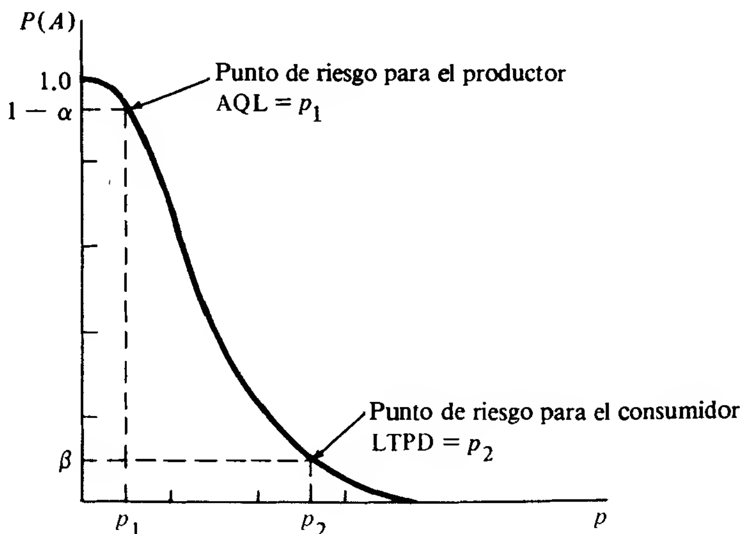


FIGURA 11.5 Curva CO para los puntos de riesgo especificados para el productor y el consumidor

este contexto, p_1 recibe el nombre de *nivel aceptable de calidad* (NAC), y p_2 el de *tolerancia de la proporción de defectuosos en el lote* (TPDL). La práctica usual ha sido la de escoger la probabilidad de aceptación $P(A) = 1 - \alpha$ en NAC cercano al punto 0.95 de la curva, y la probabilidad de aceptación $P(A) = \beta$ en TPDL cercano al punto 0.10 sobre la curva. Entonces, el 95% de los lotes que provienen de un proceso cuya proporción de artículos defectuosos se encuentra en NAC, o por encima de éste, se aceptará, mientras que sólo el 10% de los que provienen de un proceso cuya proporción de artículos defectuosos se encuentra en TPDL o más, será aceptada.

11.3.1 El desarrollo de planes de muestreo sencillos para riesgos estipulados del productor y del consumidor

Se examinará un procedimiento para obtener planes de muestreo sencillos para valores especificados de los riesgos del productor y del consumidor. La esencia del procedimiento está en determinar el tamaño de la muestra n y el número de aceptación c, dadas las probabilidades de aceptación en el NAC y el TPDL. Por ejemplo, supóngase que se desea un plan sencillo de muestreo para el que la curva característica de operación pasa a través de un riesgo del productor $\alpha = 0.05$ en un NAC de 0.01, y de un riesgo del consumidor $\beta = 0.1$ en un TPDL de 0.05. De esta forma, las probabilidades de aceptación al NAC = 0.01 y TPDL = 0.05 son 0.95 y 0.1, respectivamente.

Supóngase que las condiciones son tales, que la distribución de Poisson proporcionará una aproximación adecuada. Sea X la variable aleatoria que representa el número de artículos defectuosos en una muestra de tamaño n. Entonces para el riesgo del productor, se desea obtener n y c, tales que

$$P(A) \equiv P(X \leq c) = \sum_{x=0}^{c} \frac{\lambda^x \exp(-\lambda)}{x!} = 1 - \alpha, \tag{11.12}$$

en donde $\lambda = np_1$. De manera similar, para el riesgo del consumidor, se desea obtener n y c, tales que

$$P(A) \equiv P(X \leq c) = \sum_{x=0}^{c} \frac{\lambda^x \exp(-\lambda)}{x!} = \beta, \tag{11.13}$$

en donde ahora $\lambda = np_2$. Dado que se conocen los valores de α, β, p_1 y p_2, el procesamiento se reduce a la solución simultánea de (11.12) y (11.13) para n y c. No existe ningún método directo para resolver estas dos ecuaciones; en otras palabras, es virtualmente imposible determinar un plan de muestreo cuya curva CO pasa en forma exacta a través de dos puntos $(p_1, 1 - \alpha)$ y (p_2, β) debido a que los valores de n y c deben ser números enteros. Lo que se hace en forma general, es obtener cuatro planes, dos de los cuales tendrán el valor dado de α pero diferirán muy poco para el valor de β, mientras que los otros dos tendrán el valor de β dado, pero diferirán muy poco del valor de α.

Dados $\alpha = 0.05$, $\beta = 0.1$, $p_1 = 0.01$, y $p_2 = 0.05$, el procedimiento es el siguiente: sea $\lambda_1 = np_1$ y $\lambda_2 = np_2$ y fórmese el cociente de λ_2 a λ_1. Para el ejem-

plo se observa que el valor de éste es 5. En forma ideal, lo que se busca es obtener el valor de c cuando λ_2/λ_1 es exactamente 5. Dado que no es probable tener este valor de manera precisa, lo que se desea es determinar los dos valores de c que se encuentran relacionados con el valor de 5. Los anterior puede lograrse si se inicia con $c = 0$ y se interpola, para encontrar valores λ_1, tales que $P(A) = 1 - \alpha$, y para λ_2, tales que $P(A) \doteq \beta$, mediante el empleo de la distribución acumulativa de Poisson (tabla B del apéndice). Entonces se aumenta el valor de c, y se continúa el proceso hasta que se encuentren los valores de c que estén relacionados con el cociente deseado. Los tamaños correspondientes de las muestras se obtienen, primero, al fijar la probabilidad de aceptación del riesgo del productor dado, y después al hacer lo mismo para el riesgo del consumidor, este procedimiento dará como resultado cuatro planes de muestreo diferentes.

Dado que $P(A) = 0.95$ y $c = 0$, se obtiene que $\lambda_1 = 0.05$. De manera similar, para $P(A) = 0.1$ y $c = 0$, λ_2 tiene un valor de 2.30, y para el cociente $\lambda_2/\lambda_1 = 46$. Ahora, para $P(A) = 0.95$ y $c = 1$, $\lambda_1 = 0.36$, y para $P(A) = 0.10$, $\lambda_2 = 3.9$. De esta forma $\lambda_2/\lambda_1 = 10.83$. El proceso continúa y se obtienen los resultados que se muestran en la tabla 11.4. Los dos valores de c que se relacionan con el cociente ideal de 5 son 2 y 3.

Para obtener n, supóngase que se mantiene el riesgo del productor en $\alpha = 0.05$. Entonces para $c = 2$, $np_1 = 0.82$; pero $p_1 = 0.01$ y $n = 82$. Para el plan $n = 82$ y $c = 2$, la probabilidad de aceptar a un nivel TPDL = 0.05 se obtiene mediante $\lambda_2 \doteq (82)(0.05) = 4.1$. De cuerdo con lo anterior $P(A) \equiv P(X \leq 2) = 0.2238$.

Si se fija el riesgo del consumidor en $\beta = 0.1$, entonces para $c = 2$, $np_2 = 5.32$, y $n = 107$. Como resultado se tiene que $\lambda_1 = (107)(0.01) = 1.07$, y la probabilidad de aceptar en un NAC = 0.01 es $P(A) \equiv P(X \leq 2) \approx 0.91$. Se pueden establecer los otros dos planes si se repite el proceso anterior con $c = 3$. En la tabla 11.5 se resumen los cuatro planes; de éstos, el que parece tener la menor importancia con respecto al riesgo especificado del consumidor es $n = 82$ y $c = 2$. Los otros tres, en especial los últimos dos, se encuentran cercanos a los riesgos especificados, tanto del productor como del consumidor. La decisión final sobre cuál adoptar se toma con base en las circunstancias de la situación.

11.3.2 Muestreo para aceptación por variables

La mayoría de los planes de muestreo para aceptación se llevan a cabo por características, debido a dos razones fundamentales: la inspección por características es

TABLA 11.4 Determinación de los valores de c que se encuentran relacionados con $\lambda_2/\lambda_1 = 5$.

Número de aceptación c	*Valor de $\lambda_1 = np_1$ para $P(A) = 0.95$*	*Valor de $\lambda_2 = np_2$ para $P(A) = 0.1$*	λ_2/λ_1
0	0.05	2.30	46.00
1	0.36	3.90	10.83
2	0.82	5.32	6.49
3	1.37	6.68	4.88

TABLA 11.5 Cuatro planes de muestreo para $\alpha = 0.05$, $\beta = 0.1$, NAC = 0.01, y TPDL = 0.05.

Plan de muestreo	*Probabilidad de aceptación para* NAC = 0.01	*Probabilidad de aceptación para* TPDL = 0.05
$n = 82$, $c = 2$	0.95	0.2238
$n = 107$, $c = 2$	0.91	0.10
$n = 137$, $c = 3$	0.95	0.09
$n = 134$, $c = 3$	0.95	0.10

muy económica y muchas de las características de calidad sólo son observables como atributos. Sin embargo, en algunos casos puede hacerse una medición física de la calidad de un producto dado. Cuando la aceptación se hace con base en mediciones físicas se dice que el muestreo se lleva a cabo por variables. Cuando éste es posible, se convierte en el tipo de muestreo más popular, ya que una medición física es probable que proporcione mucho más información útil con respecto a la calidad de un producto que la dada por característica. Además, pueden obtenerse curvas CO más pronunciadas para el mismo tamaño de la muestra. La inspección por variables en general es más costosa que la inspección por características, debido a que, principalmente, tiene que aplicarse el criterio de aceptación por separado para cada medición de calidad cuando se muestrea por variables.

En el caso sencillo en el que la aceptación de un lote se hace con base en las medias de la muestra, se supone que la medición de la calidad es una variable aleatoria normalmente distribuida y con varianza conocida. Sean α el riesgo del productor y μ_α el promedio del lote para el que la probabilidad de aceptación es $1 - \alpha$. En forma similar, sea β el riesgo del consumidor y μ_β el promedio del lote para el cual la probabilidad de aceptación es β. Es decir, si el lote tiene una media μ_α, se desea aceptar el lote con una probabilidad $1 - \alpha$, y si éste tiene una media μ_β ($\mu_\alpha > \mu_\beta$) se desea aceptar el lote con una probabilidad β. Dados α, β, μ_α, y μ_β, el plan de muestreo por variables es una muestra de tamaño n y un valor de aceptación $\bar{x}_a$, tales que, cuando el valor observado de la media de la muestra $\bar{X}$ es mayor que $\bar{x}_a$, el lote será aceptado.

Para obtener $\bar{x}_a$ y n, considérese lo siguiente. Para el riesgo del productor

$$P(\bar{X} \leqslant \bar{x}_a) = \alpha$$

o

$$P\left(Z \leqslant \frac{\bar{x}_a - \mu_\alpha}{\sigma/\sqrt{n}}\right) = \alpha,$$

en donde

$$\frac{\bar{x}_a - \mu_\alpha}{\sigma/\sqrt{n}} = z_\alpha. \tag{11.14}$$

Para el riesgo del consumidor

$$P(\overline{X} > \bar{x}_a) = \beta$$

o

$$P\left(Z > \frac{\bar{x}_a - \mu_\beta}{\sigma/\sqrt{n}}\right) = \beta,$$

en donde

$$\frac{\bar{x}_a - \mu_\beta}{\sigma/\sqrt{n}} = z_{1-\beta}. \qquad (11.15)$$

Las ecuaciones dadas por (11.14) y (11.15) dependen de las incógnitas $\bar{x}_a$ y n. Al resolver (11.14) y (11.15) para $\bar{x}_a$, se tiene

$$\bar{x}_a = \frac{\sigma}{\sqrt{n}} z_\alpha + \mu_\alpha \qquad (11.16)$$

y

$$\bar{x}_a = \frac{\sigma}{\sqrt{n}} z_{1-\beta} + \mu_\beta. \qquad (11.17)$$

Al igualar (11.16) y (11.17) y resolver para n, se tiene

$$n = \left[\frac{\sigma(z_{1-\beta} - z_\alpha)}{\mu_\alpha - \mu_\beta}\right]^2 \qquad (11.18)$$

Cuando se emplea (11.18) para obtener el tamaño de la muestra, el valor de aceptación $\bar{x}_a$ se obtiene, ya sea de (11.16) o de (11.17).

Ejemplo 11.3 La compañía constructora de un gran edificio de oficinas se interesa en la resistencia a la compresión del concreto que se empleará en la construcción del edificio. El proceso a través del cual se fabrica el concreto con una resistencia promedio de 350 kilogramos por centímetro cuadrado es bueno. El concreto adquirido en este proceso debe aceptarse el 95% de las veces. Un proceso que ofrece una resistencia promedio de 347 kilogramos por centímetro cuadrado no es efectivo, y al ser adquirido será rechazado el 90% de las veces. Si el fabricante de cemento asegura a la compañía que la desviación estándar de su proceso no es mayor de 5 kilogramos por centímetro cuadrado, ¿cuántas muestras de concreto debe inspeccionar el contratista con respecto a su resistencia, y cuál debe ser el valor de aceptación para la media de la muestra bajo las condiciones dadas? Supóngase que la resistencia a la compresión del concreto se encuentra normalmente distribuida.

Los riesgos del productor y del consumidor están dados como $\alpha = 0.05$ para $\mu_\alpha = 350$ y $\beta = 0.10$ para $\mu_\beta = 347$, respectivamente. Para $\alpha = 0.05$ y $1 - \beta = 0.9$, los valores cuantiles normales estandarizados correspondientes son $z_{0.05} = -1.645$ y $z_{0.9} = 1.282$. Entonces, mediante el empleo de (11.18), el tamaño necesario de la muestra es

$$n = \left[\frac{5(1.282 + 1.645)}{350 - 347}\right]^2 = 24.$$

Para el riesgo del productor (11.16)

$$\bar{x}_a \doteq \frac{5}{\sqrt{24}}(-1.645) + 350 = 348.32,$$

y para el del consumidor (11.17)

$$\bar{x}_a = \frac{5}{\sqrt{24}}(1.282) + 347 = 348.31.$$

Para $\bar{x}_a = 348.32$, el plan de muestreo consiste en probar la resistencia de 24 muestras de concreto provenientes del proceso v aceptar el concreto siempre que la resistencia promedio sea mayor de 348.32 kilogramos por centímetro cuadrado.

11.3.3 Sistemas de planes de muestreo

Desde la Segunda Guerra Mundial, los planes de muestreo para aceptación se han convertido en procedimientos estándar para asegurar la calidad de los productos manufacturados y con este propósito se ha desarrollado una gran variedad de sistemas de planes de muestreo para aceptación. Tres de los sistemas más empleados son MIL-STD-105D*, MIL-STD-414, y el Dodge-Romig Sampling Inspection Tables. En las referencias [4], [5] y [1] se encuentra información detallada de estos sistemas. Los primeros dos fueron desarrollados por el Departamento de la Defensa y se aplican bajo un procedimiento de inspección no verificable. MIL-STD-105D contiene planes para el muestreo por características y MIL- STD-414 para el muestreo por variables. Los planes de muestreo Dodge-Romig se basan en un programa de inspección con verificación; éstos suponen un porcentaje de unidades defectuosas del proceso conocido, y los planes de muestreo sencillos se encuentran *indexados* por TPDL para riesgo del consumidor de 0.10. Estos tres sistemas se encuentran descritos en [3].

Referencias

1. H. F. Dodge and H. G. Romig. *Sampling inspection tables — Single and double sampling*, 2nd ed. Wiley, New York, 1959.
2. A. J. Duncan, *Quality control and industrial statistics*, 4th ed., Richard D. Irwin, Homewood, Ill., 1974.
3. E. L. Grant and R. S. Leavenworth, *Statistical quality control*, 4th ed., McGraw-Hill, New York, 1972.
4. *Military standard 105D, Sampling procedures and tables for inspection by attributes*, Superintendent of Documents, Government Printing Office, Washington, D.C., 1963.

* Fuera de Estados Unidos el sistema se conoce como ABC-STD-105D.

5. *Military standard 414, Sampling procedures and tables for inspection by variables for percent defective*, Superintendent of Documents, Government Printing Office, Washington, D.C., 1957.

Ejercicios

11.1. El consejo estatal formado para controlar la calidad del agua selecciona cada semana cinco muestras de agua de una fuente de abastecimiento y determina la concentración promedio de una sustancia tóxica. Los siguientes datos son las cantidades promedio en partes por millón durante 12 semanas.

Semana	1	2	3	4	5	6	7	8	9	10	11	12
Media de la muestra	5.2	4.9	5.5	5.4	4.8	4.6	5.5	4.7	5.1	4.5	5.8	5.6

a) Si los valores de la concentración promedio y de la desviación estándar son 5 y 0.5 ppm, respectivamente, obténganse los límites de control 3σ para la concentración promedio. Para este periodo, ¿existió alguna razón para alarmarse?

b) Si se considera como peligrosa una concentración de 6 ppm, ¿que tan probable es tener un resultado como el anterior, con base en cinco muestras de agua, si la concentración real promedio es de 5 ppm?

c) Mediante el uso de los límites de control de la parte *a*, ¿cuál es la probabilidad de detectar un desplazamiento en el valor de la concentración media de 5 ppm a 5.25 ppm?

11.2. Mediante el empleo de la información proporcionada en el ejercicio 11.1, obténganse los límites de control 3σ para la desviación estándar de la muestra.

11.3. Los siguientes datos son las tensiones de ruptura promedio de seis muestras de metal tomadas en forma periódica:

Muestra	1	2	3	4	5	6	7	8	9	10
Media de la muestra	498.6	508.3	484.6	505.7	491.7	495.4	482.6	515.2	510.8	503.7

Se sabe que los valores de la tensión de ruptura promedio y de la desviación estándar son 500 y 20 libras, respectivamente.

a) Obténganse los límites de control 3σ para la tensión de ruptura media de la muestra y hágase una gráfica de la tabla de control. ¿Existe alguna media muestral que se encuentre fuera de los límites de control?

b) Obténgase la probabilidad de no detectar un corrimiento en el valor real de la tensión de ruptura promedio de 500 a 494 libras.

c) Obténganse los límites de control 3σ para la desviación estándar muestral.

11.4. Los datos que se encuentran en la tabla 11.6 consisten en 20 muestras, cada una con cuatro observaciones, de los diámetros de cojinetes producidos por un proceso de manufactura.

a) Constrúyanse los límites tentativos 3σ para las tablas de control $\overline{X}$ y S.

b) Si se detecta que el proceso no se encuentra bajo control, con base en alguna muestra, recalcúlense los límites tentativos.

TABLA 11.6 Datos de la muestra para el ejercicio 11.4

Número de la muestra	*Valores de la muestra (en centímetros)*			
1	4.01	4.03	3.98	4.04
2	3.97	3.99	3.99	4.02
3	4.06	4.05	3.97	4.02
4	3.96	3.98	4.07	4.03
5	3.98	3.99	3.99	4.00
6	4.01	4.02	3.96	3.99
7	3.95	3.98	4.02	4.03
8	4.03	4.00	3.96	4.04
9	4.07	3.96	3.98	4.05
10	3.98	3.97	4.02	4.04
11	3.92	4.03	4.05	3.99
12	3.97	4.05	4.04	4.01
13	4.04	4.04	3.96	3.99
14	4.03	4.00	4.02	4.05
15	3.95	3.96	3.95	4.02
16	4.05	4.09	4.07	4.02
17	3.98	4.06	4.04	4.03
18	4.01	4.02	4.00	3.97
19	4.02	4.01	4.05	3.99
20	3.99	3.99	4.01	4.00

11.5. Las tablas de control $\overline{X}$ y S de un proceso de llenado de recipientes se conservan por algún tiempo. Con base en 25 muestras periódicas, cada una con cinco recipientes, se obtiene que $\overline{X} = 400.2$ g y $\overline{S} = 15.3$ g.

a) Si se supone que el proceso de llenado se encuentra bajo control ¿cuáles son los límites de control de la media y la desviación estándar muestral?
b) Obténgase un estimado de la desviación estándar del proceso.

11.6. En el ejercicio 11.5, supóngase que cada muestra contenía seis recipientes. ¿Cómo puede afectar este cambio a las respuestas de las partes *a* y *b*?

11.7. En un proceso de manufactura, cada día se seleccionan al azar 100 unidades y se envían para su inspección. Los siguientes datos son el número de unidades defectuosas en la muestra durante 25 días.

Día	1	2	3	4	5	6	7	8	9	10	11	12	13
Número de unidades defectuosas	2	1	4	3	2	2	5	3	4	2	1	5	2

Día	14	15	16	17	18	19	20	21	22	23	24	25
Número de unidades defectuosas	3	2	1	0	6	4	5	2	1	8	3	2

a) Con base en esta información, obténgase una tabla *p*.
b) Revísense los límites de control si algún día el proceso se juzgó como fuera de control.

c) Si se supone que el proceso se encuentra bajo control con un porcentaje de unidades defectuosas, igual al obtenido en la parte *b*, ¿cuál es la probabilidad de que, en un día determinado el proceso se considere como fuera de control?

11.8. Se supone que el porcentaje de unidades defectuosas para un proceso de manufactura es de 4%. El proceso se vigila diariamente mediante la toma de muestras de $n = 80$ unidades. Éste se detiene cada vez que se encuentran cinco o más unidades defectuosas en la muestra. Si el verdadero porcentaje de unidades defectuosas es de 5.5%, ¿cuál es la probabilidad de detener el proceso?

11.9. Supóngase que la calidad de un lote muy grande es de sólo 5% de unidades defectuosas. Un plan de muestreo para aceptación requiere una muestra de 40 unidades y un número de aceptación igual a 2 unidades.

a) ¿Cuál es la probabilidad de que el lote sea aceptado?
b) Si la calidad real del lote es de 6.25% de unidades defectuosas, ¿cuál es la probabilidad de que el lote sea aceptado?

11.10. Para el ejercicio 11.9, supóngase que el tamaño de la muestra es de $n = 80$ unidades y el número de aceptación es igual a cuatro unidades. ¿Cómo afectarán estos cambios a las respuestas de las partes *a* y *b*?

11.11. La calidad de un lote de $N = 20$ unidades es del 10% defectuosas. Si se toma una muestra aleatoria de cinco unidades y no se encuentra ninguna defectuosa se aceptará el lote. ¿Cuál es la probabilidad de aceptar el lote?

11.12. Hágase una gráfica de las curvas características de operación para los planes de muestreo $n = 25$, $c = 1$ y $n = 50$, $c = 2$. Compárense las curvas características de operación.

11.13. Para el plan de muestreo $n = 25$, $c = 1$, empléese la curva CO para obtener el TPDL para un riesgo del consumidor de 0.05.

11.14. Para el plan de muestreo $n = 50$, $c = 2$, empléese la curva CO para obtener el NAC para un riesgo del productor de 0.05.

11.15. Obténganse los cuatro planes de muestreo que relacionarán los riesgos del productor y del consumidor de $\alpha = 0.05$ para NAC $= 0.02$ y $\beta = 0.1$ para TPDL $= 0.08$, respectivamente.

11.16. Obténganse los cuatro planes de muestreo que relacionarán los riesgos del productor y del consumidor de $\alpha = 0.10$ para NAC $= 0.01$ y $\beta = 0.1$ para TPDL $= 0.05$.

11.17. En muchas ocasiones se emplea un plan de muestreo doble para el muestreo de aceptación; este plan requiere una muestra aleatoria de n_1 unidades de un lote de N unidades. Si el número de unidades defectuosas no es mayor que c_1, el lote se acepta; si se encuentra una cantidad de unidades defectuosas $c_2 > c_1$ el lote se rechaza. Si el número de unidades defectuosas en la primera muestra es mayor que c_1, pero menor que c_2, se toma otra muestra aleatoria de tamaño n_2. El lote se acepto si el número de unidades defectuosas en ambas muestras no es mayor que c_2; de otra forma el lote se rechaza. Mediante el empleo de este procedimiento determínense las siguientes probabilidades para el doble plan de muestreo $N = 5000$, $n_1 = 50$, $n_2 = 80$, $c_1 = 0$, $c_2 = 3$ si la calidad del lote es de 2% de unidades defectuosas.

a) La probabilidad de aceptar el lote con base en la primera muestra.

b) La probabilidad de rechazar el lote con base en la primera muestra.
c) La probabilidad de aceptar el lote después de tomar la segunda muestra.
d) La probabilidad de rechazar el lote después de tomar la segunda muestra.

11.18. Una agencia estatal se encarga de vigilar el nivel de concentración de cierto contaminante químico, el cual ha sido derramado en grandes cantidades en uno de los ríos más grandes del estado. La agencia debe decidir en forma periódica cuándo el nivel de concentración se encuentra entre límites seguros para permitir la pesca con fines comerciales. La agencia desea obtener un plan de muestreo por variables de tal manera que cuando el nivel de concentración promedio real sea de 5.6 ppm decidirá el 95% de las veces que la pesca continúe. Pero desea prohibir la pesca el 99% de las veces que se observe una concentración hasta de 6.0 ppm. Si la desviación estándar no es mayor de una parte por millón, determínese el plan de muestreo. Supóngase que la concentración de este contaminante se encuentra normalmente distribuida.

11.19. Un comprador de grandes cantidades de hilo desea desarrollar un plan de muestreo por variables para la tensión de ruptura del hilo. El hilo será aceptado por el comprador si su tensión de ruptura es mayor de 60 libras. Si se sabe que la desviación estándar del hilo es de 8 libras y dados $\alpha = 0.05$, $\beta = 0.05$, NAC = 0.05 y TPDL = 0.1, obténgase el plan de muestreo. Supóngase que la tensión del hilo se encuentra normalmente distribuida.

CAPÍTULO DOCE

Diseño y análisis de experimentos estadísticos

12.1 Introducción

En las secciones 9.6.3 y 9.6.4 se introdujeron algunas ideas básicas con respecto a la planeación y adquisición de datos experimentales, con el propósito de alcanzar el máximo beneficio de la aplicación de la inferencia estadística. En este capítulo se estudiará la noción de experimentos diseñados estadísticamente y se extenderán algunos de los métodos del capítulo 9 mediante la introducción de una técnica estadística importante conocida como análisis de varianza.

12.2 Experimentos estadísticos

Para cualquier fenómeno en el que existe la incertidumbre, el procedimiento apropiado para investigarlo es experimentar con él, de manera que puedan identificarse las características de interés. Por ejemplo, supóngase que se desea identificar el comportamiento óptimo de un sistema con respecto a su funcionamiento y costo en distintas condiciones; entonces debe pensarse en un experimento como medio para que el sistema sea observado bajo las condiciones de interés, de tal manera que su comportamiento pueda conocerse.

El elemento más importante de un experimento, y que muchas veces se subestima, es la formulación del problema por resolver. No puede esperarse una oportunidad de éxito razonable sin alguna dirección con respecto al propósito del experimento. Una vez que éste se define, es necesario identificar la variable por medir o *respuesta* que se va a estudiar y el *factor* o *factores* potenciales que pueden influenciar la variabilidad de la respuesta. La respuesta también se conoce como *variable dependiente;* el factor recibe el nombre de *variable independiente;* se supone que este último se encuentra bajo el control del investigador. Por ejemplo, en una tienda el interés recae en el número de empleados disponible, de manera que el tiempo de espera del cliente no sea excesivo. En este caso, la respuesta es el tiempo de espera y el factor el número de empleados disponible.

Un *nivel* o *tratamiento* del factor es un valor o condición de éste bajo el cual se observará la respuesta medible. Por ejemplo, supóngase que se desea observar el tiempo de espera cuando la tienda tiene a su servicio dos, cuatro o seis empleados a la vez. Si un experimento consiste en varios factores, un tratamiento es una combinación de los niveles de cada factor; por ejemplo, si se desea estudiar el tiempo de espera como una función del número de empleados en un determinado momento del día, entonces un tratamiento es la combinación de un número particular de empleados en un momento dado del día. El proceso por medio del cual se seleccionan los tratamientos se encuentra dictado más o menos por las metas del experimento. Para experimentos preliminares, en los cuales el propósito primordial es aislar los principales factores, el investigador debe escoger mentalmente los tratamientos con una visión muy amplia, de manera que obtenga un conocimiento útil del mecanismo bajo estudio. En forma posterior, se puede conducir un experimento más preciso con el propósito de hacer hallazgos más específicos.

Una *unidad experimental* se define como el objeto (persona o cosa) que es capaz de producir una medición de la variable de respuesta después de aplicar un tratamiento dado. La selección de una unidad experimental o del tamaño de ésta descansa, de nuevo, enteramente en el experimentador. Por ejemplo, si un fabricante de focos desea comparar la duración de éstos con la de sus competidores, entonces los focos seleccionados son las unidades experimentales y el número de marcas diferentes los tratamientos. O si se tiene interés en determinar la concentración de un contaminante en un lago en función de la ubicación geográfica, entonces las localidades del lago que se seleccionan para medir la concentración del contaminante son los tratamientos y la pequeña área superficial de cada localidad, la unidad experimental.

En un ambiente de incertidumbre los experimentos son, en forma general, comparativos en el sentido de que, idealmente, miden y comparan las respuestas de unidades experimentales esencialmente idénticas, después de que éstas se exponen a los tratamientos seleccionados y aplicados por el investigador. Todos los factores externos que pueden influenciar la respuesta deben eliminarse o controlarse. Sin embargo, no siempre puede garantizarse el control de los factores externos; por ejemplo, en forma práctica, casi cualquier experimento que incluye alguna actividad financiera guardará alguna interrelación con las condiciones económicas prevalecientes que no pueden controlarse. Tal desviación del control experimental ideal necesita de la repetición del experimento en una muestra de unidades experimentales para determinar la variación aleatoria o *error experimental.* Esta es la variación extraña en la respuesta o la variación que no puede ser atribuible a un cambio de tratamiento. Por lo tanto, es posible la inferencia estadística al comparar el error experimental con las respuestas promedio que resultan de la aplicación de los diferentes tratamientos.

En algunas ciencias pueden llevarse a cabo experimentos de laboratorio ideales, pero en las ciencias socioeconómicas, las desviaciones de las condiciones experimentales ideales tienen un lugar común debido a que el medio no permite un control suficiente. Por ejemplo, puede ser interesante estudiar el efecto de un aumento en las tasas de interés (tramiento) en la actividad de construcción de casas (respuesta) por parte de los constructores (unidades experimentales). Los tratamientos no pueden aplicarse a las unidades experimentales, ni la respuesta puede medirse de acuerdo con un experimento planeado. Sólo puede registrarse la información conforme cambian las condi-

ciones en el mundo real. Aunque para un purista lo anterior no constituye un experimento, estos tipos de estudios merecen una considerable atención. Para el análisis de estos datos es más apropiado el empleo de los métodos de regresión que los que se estudiarán en este capítulo. En los capítulos 13 y 14 se examinará el análisis de regresión.

12.3 Diseños estadísticos

El proceso por medio del cual se miden las observaciones de la respuesta se centra en un diseño estadístico. En general, en los experimentos diseñados estadísticamente, las unidades experimentales deben seleccionarse en forma imparcial, así como los tratamientos asignados a éstas, mediante un proceso aleatorio, con el propósito de remover los posibles sesgos sistemáticos. Como ya se indicó en el capítulo 9, el proceso aleatorio no sólo protege contra el sesgo sistemático, sino también tiende a neutralizar los efectos de todos aquellos factores externos que no se encuentren bajo el control del investigador. Entonces las comparaciones entre los tratamientos se miden, en forma práctica, como si el efecto en la respuesta se debiera sólo a la diferencia entre los tratamientos.

En un experimento diseñado estadísticamente es de igual importancia el concepto de *repetición.* Como ya se ha notado con anterioridad, el propósito de la repetición es medir el error experimental. La magnitud de éste juega un papel muy importante en la toma de decisiones con respecto a la posibilidad de que las diferencias entre los tratamientos sean discernibles en forma estadística.

En el diseño de experimentos estadísticos, el interés primario recae en cómo asignar las unidades experimentales a los tratamientos (o viceversa), para asegurar un proceso imparcial. En este contexto surgen dos conceptos básicos: el proceso de asignación debe hacerse con base en un *diseño completamente aleatorio,* o en un *diseño en bloque completamente aleatorio.* Cualquiera de estos dos diseños puede emplearse en experimentos unifactoriales o en aquéllos en los que se desea investigar varios factores en forma simultánea. Con un diseño complementario aleatorio, la asignación de los tratamientos a cada unidad experimental se lleva a cabo en forma totalmente aleatoria y todas las unidades se suponen homogéneas. En forma general, se hace uso de un procedimiento aleatorio sencillo como la generación de números aleatorios para llevar a cabo el proceso de asignación. El uso de un diseño completamente aleatorio implica que las condiciones bajo las cuales será observada la respuesta (u otras que se encuentren bajo el control del investigador) serán las mismas a través de todo el experimento. Este tipo de diseño no debe usarse en aquellas situaciones en las que las observaciones se realizarán sobre factores potenciales como el tiempo, el espacio o efectos demográficos, a menos que éstos sean partes legítimas del experimento.

No obstante, muchas veces el investigador se da cuenta de que el experimento no se puede conducir en el mismo ambiente, debido, principalmente, a que no todas las unidades experimentales son homogéneas; por lo tanto, éstas se clasifican en *bloques* homogéneos y se asignan todos los tratamientos en forma aleatoria a las unidades de cada bloque, con lo que se crea lo que se conoce como un diseño en bloques completamente aleatorio. La palabra "completamente" indica que cada bloque contiene todos los

tratamientos, mientras que la palabra "aleatorio" significa que todos los tratamientos serán asignados, en forma aleatoria, a las unidades experimentales de cada bloque.

El investigador reconoce la necesidad de agrupar en bloques, mediante la identificación de los elementos potenciales de las unidades experimentales que no se han incluido en la definición de un tratamiento, pero que pueden causar una variación significativa en la respuesta. Muchas veces éstos guardan relación con efectos espaciales, temporales o demográficos. Por ejemplo, si las unidades experimentales son seres humanos, entonces el agrupamiento por bloques deberá hacerse tomando en cuenta sexo, edad, condiciones de salud, experiencia, etc., como lo dicta el experimento. Si éste se va a realizar en un lapso grande deberá considerarse como una variable para el agrupamiento por bloques. Si los datos experimentales se van a recolectar, ya sea en distintas localidades o en grupos, entonces éstos deberán considerarse como variables en bloque. Si se van a usar varios instrumentos para registrar los datos, se deberá considerar un agrupamiento de instrumentos por bloques, aun si éstos son del mismo modelo y con mayor razón si provienen de distintos fabricantes.

Por lo tanto, la necesidad de agrupar en bloques es evidente; entre más heterogéneas son las unidades experimentales, mayor es el error experimental y menor la oportunidad de detectar diferencias reales entre los diversos tratamientos. La razón de agrupar en bloques es tomar en cuenta, y de esta forma remover, la fuente de variación en la respuesta que no es de interés, con lo que se incrementa la sensitividad para detectar diferencias entre los tratamientos. Así, el principio general de un diseño estadístico radica en minimizar el error experimental mediante el control de las variaciones extrañas, de manera que pueda detectarse la variación sistemática en la respuesta.

12.4 Análisis de experimentos unifactoriales en un diseño completamente aleatorio

El tipo de experimento más sencillo es aquél que compara el efecto de $k \geqslant 2$ niveles de un solo factor sobre alguna variable de respuesta. Los niveles del factor son los tratamientos, y si éstos se aplican en forma aleatoria a un conjunto virtualmente homogéneo de unidades experimentales, el experimento tiene un diseño completamente aleatorio. Esta situación es una extensión natural del problema que surge cuando se comparan dos medias poblacionales en donde las variantes son desconocidas pero que se suponen iguales. La prueba t para dos muestras, la cual se estudió en el capítulo 9, se basa en un diseño completamente aleatorio.

Para $k \geqslant 2$ niveles, se desea probar la hipótesis nula

$$H_0\colon \mu_1 = \mu_2 = \cdots = \mu_k \tag{12.1}$$

contra la alternativa de que algunas de las medias de la población no son las mismas. Si es posible rechazar la hipótesis nula con base en k muestras independientes, entonces las medias de las k poblaciones no son todas iguales entre sí, o el efecto de los tratamientos sobre la respuesta es estadísticamente discernible. Si no puede rechazarse la hipótesis nula, cualquier desviación observada en la respuesta se debe sólo al error aleatorio y no a causa de un cambio en el tratamiento.

Se pueden manejar muchos problemas prácticos con un experimento unifactorial completamente aleatorio. Unos cuantos ejemplos son los siguientes: saber si tienen algún efecto sobre el consumo de energía ligeras diferencias en el aislamiento de los techos de las casas; si la media del llenado producido por máquinas en un proceso de llenado es la misma, o si los vendedores que reciben diferentes métodos de entrenamiento, incrementan su volumen de ventas en forma diferente. En estos casos, los tratamientos son el aislamiento de los techos, las diferentes máquinas y los diversos métodos de entrenamiento; las unidades experimentales son las causas seleccionadas, los recipientes llenos y los vendedores, respectivamente. En el primer caso los tratamientos son cuantitativos, ya que los distingue una escala bien definida (R). En los últimos dos casos los tratamientos son cualitativos, dado que representan cosas o sujetos diferentes y por lo tanto carecen de escalas numéricas.

La necesidad de tener unidades experimentales homogéneas esencialmente puede ilustrarse con el primer ejemplo. Si se seleccionan casas para el experimento que no sean del mismo tamaño, en ese caso no se tiene el mismo aislamiento en los techos y se tienen distintas calidades con respecto al clima, si éstas se localizan en distintas zonas geográficas; de esta forma las diferencias en el consumo de energía no se pueden atribuir sólo al aislamiento del techo. Así, para un diseño completamente aleatorio los resultados serán ambiguos, a menos que las unidades experimentales sean virtualmente homogéneas.

La técnica del *análisis de varianza* proporciona el procedimiento inferencial para probar la hipótesis nula dada por (12.1). Para desarrollar esta técnica, se analizará el problema del aislamiento. Supóngase que se tiene interés en k diferentes niveles de aislamiento en el techo, tales que para el j-ésimo nivel se observará el consumo de energía mensual del sistema de calentamiento en n_j casas diferentes pero muy similares. Las casas que se seleccionan para este experimento son homogéneas y los factores externos están controlados dentro de ciertos límites prácticos. La información de la muestra puede colocarse como se presenta en la tabla 12.1, donde la respuesta medible es el número de kilowats-hora mensuales utilizados por el sistema de calentamiento de cada casa.

TABLA 12.1 Arreglo común de los datos de la muestra de un experimento con sólo un factor completamente aleatorizado

		Tratamientos			
1	2	···	j	···	k
Y_{11}	Y_{12}	···	Y_{1j}	···	Y_{1k}
Y_{21}	Y_{22}	···	Y_{2j}	···	Y_{2k}
.	.		.		.
.	.		.		.
.	.		.		.
Y_{i1}	Y_{i2}	···	Y_{ij}	···	Y_{ik}
.	.		.		.
.	.		.		.
.	.		.		.
$Y_{n_1 1}$	$Y_{n_2 2}$	···	$Y_{n_j j}$	···	$Y_{n_k k}$

Se supone que cada nivel de aislamiento térmico en los techos representa una población a partir de la cual se obtiene una muestra; también, que las distribuciones de las poblaciones para cada nivel de aislamiento son normales con varianzas iguales. De acuerdo con lo anterior, las columnas de la tabla 12.1 representan k muestras aleatorias independientes de tamaños n_j, $j = 1, 2, \ldots, k$. Si la hipótesis nula dada por (12.1) es cierta, la observación Y_{ij} es el uso promedio de energía de los sistemas de calentamiento para todos los k niveles de aislamiento térmico y cualquier desviación del promedio se debe a un error aleatorio. Si H_0 es falsa, entonces Y_{ij} está constituida por todos los promedios, más el efecto del j-ésimo tratamiento y el error aleatorio. El promedio matemático para un experimento unifactorial completamente aleatorio es

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij} \qquad j = 1, 2, \ldots, k, \qquad (12.2)$$
$$i = 1, 2, \ldots, n_j,$$

en donde Y_{ij} es la i-ésima observación del j-ésimo tratamiento, μ es la media sobre todas las k poblaciones, τ_j es el efecto sobre la respuesta debido al j-ésimo tratamiento, y ε_{ij}* es el error experimental para la i-ésima observación bajo el j-ésimo tratamiento.

Se supone que los errores son independientes y que se encuentran normalmente distribuidos con medias cero y varianzas iguales. En otras palabras, $\varepsilon_{ij} \sim N(0, \sigma^2)$ para toda i y j. La suposición sobre los τ_j' depende de cómo considere el investigador los niveles del factor. Si el investigador está interesado en lo que le pasa a la respuesta, sólo para ciertos niveles del factor que se seleccionan de antemano, entonces $\tau_1, \tau_2, \ldots, \tau_k$ se consideran como parámetros fijos tales, que

$$\sum_{j=1}^{k} n_j \tau_j = 0.$$

Por lo tanto, el modelo dado por (12.2) se conoce como *modelo de efectos fijos* y las inferencias estadísticas con respecto a los efectos de los tratamientos pertenecen, en forma exclusiva, a los niveles seleccionados.

Por otro lado, si los niveles empleados en el experimento se seleccionaron al azar, de una población de posibles niveles, entonces $\tau_1, \tau_2, \ldots, \tau_k$ son variables aleatorias independientes que $\tau_j \sim N(0, \sigma_\tau^2)$ para toda j. En este caso, el modelo dado por (12.2) se conoce como *modelo de efectos aleatorios,* y las inferencias estadísticas con respecto a los niveles de un factor pertenecen a la población de niveles.

En general, para factores cuantitativos es deseable escoger niveles fijos del intervalo de interés, debido a que no es probable que una selección aleatoria proporcione una amplia cobertura de éste. La interpolación de los niveles fijos previamente seleccionados también es una práctica muy segura para factores cuantitativos. Cuando los factores son cualitativos como seres humanos, localidades o grupos, su selección sólo es importante cuando puede revelar algo con respecto a la variabilidad de la población.

*En lugar de emplear una letra mayúscula para las variables aleatorias ε_{ij}, se seguirá la tradición de utilizar la letra griega minúscula épsilon.

Para un modelo de efectos fijos, una hipótesis nula equivalente a (12.2) es

$$H_0: \tau_j = 0, \quad \text{para toda } j. \tag{12.3}$$

La hipótesis nula (12.3) establece que no existe ningún efecto de los tratamientos sobre la respuesta, lo que a su vez implica que las k medias de la población son iguales entre sí. Entonces se tiene como resultàdo que cada observación consiste en una media común y cualquier desviación con respecto a ésta se debe a la variación inherente dentro de cada población.

Para un modelo de efectos aleatorios, la hipótesis nula consiste en la proposición de que la varianza entre los τ_j (o los efectos del tratamiento) es cero; es decir,

$$H_0: \sigma_\tau^2 = 0. \tag{12.4}$$

Así, al suponer independencia entre los errores y tratamientos aleatorios,

$$Var(Y_{ij}) = \sigma^2 + \sigma_\tau^2.$$

Para el modelo de efectos aleatorios, el interés recae en hacer una evaluación de cuánto de la varianza en las observaciones se debe a diferencias reales en las medias de los tratamientos y cuánto se debe a errores aleatorios con respecto a estas medias.

En este capítulo el principal interés se centra en el modelo de efectos fijos, pero se incluirá el caso de efectos aleatorios cuando sea necesario. El punto de vista empleado para desarrollar la técnica del análisis de varianza será, en gran parte, intuitivo. Para un tratamiento teórico de la materia, véase [6].

12.4.1 Análisis de varianza para un modelo de efectos fijos

Sean $\mu_1, \mu_2, \ldots, \mu_k$ las medias de las k poblaciones, y sea μ la media de todas las poblaciones. Se define el efecto τ_j del j-ésimo tratamiento como la desviación de la j-ésima población media μ_j respecto a la media global μ. De esta forma,

$$\tau_j = \mu_j - \mu, \qquad j = 1, 2, \ldots, k.$$

En el mismo sentido, el error aleatorio correspondiente ε_{ij} de la observación Y_{ij} es la desviación de Y_{ij} con respecto de la j-ésima media μ_j o

$$\varepsilon_{ij} = Y_{ij} - \mu_j, \qquad j = 1, 2, \ldots, k,$$
$$i = 1, 2, \ldots, n_j.$$

De acuerdo con lo anterior, el modelo dado por (12.2) puede escribirse de la siguiente manera

$$Y_{ij} = \mu + (\mu_j - \mu) + (Y_{ij} - \mu_j),$$

o

$$Y_{ij} - \mu = (\mu_j - \mu) + (Y_{ij} - \mu_j). \tag{12.5}$$

La igualdad dada por (12.5) establece, en forma explícita, que cualquier desviación de una observación con respecto a la media global se debe a dos posibles causas: a la diferencia en el tratamiento o a un error aleatorio. Si se rechaza la hipótesis nula dada por (12.3), los datos de la muestra deben demostrar que la desviación total que se debe a la diferencia en el tratamiento es, suficientemente, más grande que la desviación causada por el error aleatorio. De esta forma, la técnica del análisis de varianza es en realidad un análisis de la variación de las medias y éste se logra mediante la participación de la variación total en las observaciones en componentes especificados por el modelo matemático. Esto permite determinar una estadística apropiada de tal manera que pueda tomarse una decisión con respecto a la hipótesis H_0: $\tau_j = 0$

Los parámetros $\mu_1, \mu_2, \ldots, \mu_k$ y μ no son conocidos, pero pueden estimarse con base en las observaciones de las k muestras aleatorias. Para la información de la muestra dada en la tabla 12.1 se define lo siguiente:

$$T_{.j} = \sum_{i=1}^{n_j} Y_{ij}, \qquad j = 1, 2, \ldots, k,$$

$$\overline{Y}_{.j} = T_{.j}/n_j, \qquad j = 1, 2, \ldots, k,$$

$$T_{..} = \sum_{j=1}^{k} T_{.j},$$

$$N = \sum_{j=1}^{k} n_j,$$

$$\overline{Y}_{..} = T_{..}/N.$$

De nuevo, se emplea la notación de punto para indicar que la suma se lleva a cabo sobre el correspondiente subíndice. En particular, $T_{.j}$ es la suma de las n_j observaciones en el j-ésimo tratamiento, $\overline{Y}_{.j}$ es la media de la muestra del j-ésimo tratamiento, $T_{..}$ es la suma de todas las N observaciones y $\overline{Y}_{..}$ es la media de la muestra de todas las observaciones.

Al sustituir las estadísticas $\overline{Y}_{.j}$ y $\overline{Y}_{..}$ en (12.5) para los parámetros μ_j y μ, respectivamente, se obtiene la correspondiente igualdad en la muestra

$$Y_{ij} - \overline{Y}_{..} = (\overline{Y}_{.j} - \overline{Y}_{..}) + (Y_{ij} - \overline{Y}_{.j}). \qquad (12.6)$$

La esencia de la identidad de la muestra (12.6) es la división de la desviación de una observación Y_{ij} del promedio de la muestra total $\overline{Y}$ en dos componentes la desviación de la media de la muestra del tratamiento $\overline{Y}_{.j}$ de $\overline{Y}_{..}$, y la desviación de Y_{ij} de su propia media de tratamiento $\overline{Y}_{.j}$. De acuerdo con lo anterior, puede argumentarse en forma lógica que entre mayor sea la desviación entre $\overline{Y}_{.j}$ y $\overline{Y}_{..}$, se tiene más inclinación a rechazar la hipótesis nula dada por (12.3).

Para determinar una estadística de prueba apropiada, supóngase que se toma el cuadrado de ambos miembros de (12.6) y se suman sobre todos los i y j. De esta

forma,

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\overline{Y}_{..})^2=\sum_{j=1}^{k}\sum_{i=1}^{n_j}(\overline{Y}_{.j}-\overline{Y}_{..})^2+\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\overline{Y}_{.j})^2$$
$$+\,2\sum_{j=1}^{k}\sum_{i=1}^{n_j}(\overline{Y}_{.j}-\overline{Y}_{..})(Y_{ij}-\overline{Y}_{.j}). \qquad (12.7)$$

Pero

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(\overline{Y}_{.j}-\overline{Y}_{..})(Y_{ij}-\overline{Y}_{.j})=\sum_{j=1}^{k}(\overline{Y}_{.j}-\overline{Y}_{..})\left[\sum_{i=1}^{n_j}(Y_{ij}-\overline{Y}_{.j})\right]$$
$$=\sum_{j=1}^{k}(\overline{Y}_{.j}-\overline{Y}_{..})\left[\sum_{i=1}^{n_j}Y_{ij}-n_j\overline{Y}_{.j}\right]$$
$$=0,$$

dado que $\Sigma_{i=1}^{n_j} Y_{ij} = T_{.j} = n_j\overline{Y}_{.j}$.

Como resultado se tiene que la ecuación

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\overline{Y}_{..})^2=\sum_{j=1}^{k}\sum_{i=1}^{n_j}(\overline{Y}_{.j}-\overline{Y}_{..})^2+\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\overline{Y}_{.j})^2 \qquad (12.8)$$

establece que la suma total de los cuadrados de las desviaciones con respecto a la media global se descompone en la suma de los cuadrados de las desviaciones de las medias de los tratamientos en relación con la media global, y la suma de los cuadrados de las desviaciones de las observaciones con respecto a sus propias medias de tratamiento. La expresión (12.8) se conoce como la ecuación fundamental del análisis de varianza. El término en el lado izquierdo de (12.8) es la *suma total de cuadrados* y se denota por *STC.* El término en medio de (12.8) es la *suma de los cuadrados de los tratamientos* y se denota por *SCTR*. El último término es la *suma de los cuadrados de los errores,* denotada por *SCE.* Por lo tanto,

$$STC = SCTR + SCE \qquad (12.9)$$

SCE mide la cantidad de variación en las observaciones debida a un error aleatorio. Si todas las observaciones que se encuentran dentro de un mismo tratamiento son las mismas, y si este hecho es cierto para todos los k tratamientos, entonces $SCE = 0$. De acuerdo con lo anterior, entre más grande es *SCE*, mayor es la variación en las observaciones que puede atribuirse a un error aleatorio. *SCTR* mide la extensión de la variación, en las observaciones, que se debe a las diferencias entre los tratamientos. Si todas las medias de los tratamientos son iguales entre sí, entonces $SCTR = 0$. De esta forma, entre más grande es el valor de *SCTR,* mayor es la diferencia que existe entre las medias de los tratamientos y la media global.

Puede demostrarse que bajo la hipótesis nula H_0: $\tau_j = 0$ y la suposición de que $\varepsilon_{ij} \sim N(0, \sigma^2)$, SCTR/σ^2 y SCE/σ^2 son dos variables aleatorias independientes con una distribución chi-cuadrada. Los grados de libertad se obtienen al separar la suma

total de cuadros. *STC* tiene $N - 1$ grados de libertad debido a que se pierde un grado de libertad al ser necesario que la suma de las desviaciones $(Y_{ij} - \overline{Y}_{..})$ para toda k y j sea cero. La suma de los cuadrados de los tratamientos tiene $k - 1$ grados de libertad debido a que se impone la restricción $\Sigma_{j=1}^{k} n_j(\overline{Y}_{.j} - \overline{Y}_{..}) = 0$ para las k desviaciones $(\overline{Y}_{.j} - \overline{Y}_{..})$. Esta restricción surge del hecho de que $\Sigma_{j=1}^{k} n_j\tau_j = 0$. Entonces, con base en (12.9), el número de grados de libertad para *SCE* será igual a la diferencia entre el número de grados de libertad para *STC* y *SCTR*,

$$\begin{aligned}\text{gl(SCE)} &= \text{gl(STC)} - \text{gl(SCTR)}\\ &= N - 1 - (k - 1)\\ &= N - k.\end{aligned}$$

Una suma de cuadrados dividido entre sus grados de libertad da origen a lo que se conoce como *cuadrado medio*. De acuerdo con lo anterior, el cuadrado medio del tratamiento es

$$\text{CMTR} = \text{SCTR}/(k - 1),$$

y el cuadrado medio del error es

$$\text{CME} = \text{SCE}/(N - k).$$

Ahora se puede argumentar que, dado que SCTR/σ^2 y SCE/σ^2 son dos variables aleatorias independientes chi-cuadrada con $k - 1$ y $N - k$ grados de libertad, respectivamente, entonces el cociente de las medias cuadráticas de la sección 7.8 tiene una distribución F con $k - 1$ y $N - k$ grados de libertad. Este cociente es la estadística apropiada para probar la hipótesis nula

$$H_0 : \tau_j = 0.$$

Lo anterior puede verificarse al examinar los valores esperados de los cuadrados medios. Puede demostrarse que

$$E(\text{CME}) = \sigma^2$$

y

$$E(\text{CMTR}) = \sigma^2 + \frac{\sum_{j=1}^{k} n_j\tau_j^2}{k - 1},$$

en donde σ^2 es la varianza común de los errores. Como resultado se tiene que el cuadrado medio del error es un estimador no sesgado de σ^2 sin importar si la hipótesis nula es cierta. Por otro lado, si H_0 es cierta, $\tau_j = 0$ para toda j, y $\Sigma n_j\tau_j^2 = 0$ Entonces $E(\text{CMTR}) = \sigma^2$; es decir, bajo H_0, tanto *CME* como *CMTR* son estimadores no sesgados de la varianza del error. Pero si la hipótesis nula no es de cierta, *CMTR* tiende generalmente a ser mayor que *CME*, dado que el término $\Sigma n_j\tau_j^2$ será positivo. En otras palabras, entre más grande sea la diferencia entre las medias de

los tratamientos y la media global, mayor será *CMTR*. Pero una ocurrencia de este tipo sugiere que las medias de los k tratamientos no son todas iguales entre sí y de esta forma debe rechazarse la hipótesis nula. De acuerdo con lo anterior, la hipótesis nula será rechazada cuando el valor del cociente.

$$F = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2/(k-1)}{\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}_{\cdot j})^2/(N-k)} \qquad (12.10)$$

se encuentre dentro de una región crítica superior de tamaño α.

El análisis anterior constituye la técnica del análisis de varianza para un experimento con sólo un factor completamente aleatorizado. Las fuentes de variación, grados de libertad, sumas de cuadrados, cuadrados medios, y el cociente F juntos, constituyen lo que se conoce como tabla de análisis de varianza *(ANOVA)* que se presenta en la tabla 12.2.

Dadas las verificaciones y_{ij}, $j = 1, 2, \ldots, k$, $i = 1, 2, \ldots, n_j$, el cálculo de las cantidades que aparecen en la tabla 12.2 puede hacerse en forma fácil mediante el empleo de cualquier paquete estadístico estándar para computadora. Para llevar a cabo el cálculo a mano, las sumas de los cuadrados pueden calcularse mediante el empleo de fórmulas algebraicamente equivalentes, pero desde un punto de vista de computación, más convenientes

$$\text{STC} = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_{\cdot\cdot})^2 = \sum_{j=1}^{k}\sum_{i=1}^{n_j} y_{ij}^2 - \frac{T_{\cdot\cdot}^2}{N},$$

$$\text{SCTR} = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2 = \sum_{j=1}^{k}\frac{T_{\cdot j}^2}{n_j} - \frac{T_{\cdot\cdot}^2}{N},$$

$$\text{SCE} = \text{STC} - \text{SCTR}$$

Debe notarse que la hipótesis nula H_0: $\mu_1 = \mu_2$ para el caso de dos muestras también puede manejarse con el método del análisis de varianza. En el capítulo 13 se mostrará la relación que existe entre las estadísticas F y t de Student para $k = 2$.

TABLA 12.2 Tabla de análisis de varianza para un experimento con sólo un factor completamente aleatorio

Fuente de variación	gl	SC	CM	*Estadística F*
Tratamientos	$k - 1$	$\Sigma\Sigma(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2$	$\Sigma\Sigma(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2/(k-1)$	$F = \dfrac{\Sigma\Sigma(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2/(k-1)}{\Sigma\Sigma(Y_{ij} - \bar{Y}_{\cdot j})^2/(N-k)}$
Error	$N - k$	$\Sigma\Sigma(Y_{ij} - \bar{Y}_{\cdot j})^2$	$\Sigma\Sigma(Y_{ij} - \bar{Y}_{\cdot j})^2/(N-k)$	
Total	$N - 1$	$\Sigma\Sigma(Y_{ij} - \bar{Y}_{\cdot\cdot})^2$		

TABLA 12.3 Calor empleado para cinco niveles de aislamiento

Espesor del aislamiento del techo (pulgadas)				
4	6	8	10	12
14.4	14.5	13.8	13.0	13.1
14.8	14.1	14.1	13.4	12.8
15.2	14.6	13.7	13.2	12.9
14.3	14.2	13.6		13.2
14.6		14.0		13.3
				12.7

Ejemplo 12.1 Los datos que figuran en la tabla 12.3 son los resultados de un diseño completamente aleatorizado para el cual la respuesta son los kilowats hora, empleados por los sistemas de calentamiento (en cientos de kilowats hora) para casas muy similares en un mes dado, como función de cinco niveles de aislamiento térmico (en pulgadas). Con base en esta información, ¿existe alguna razón para creer que por lo menos algunos de los consumos de energía promedio para los cinco niveles de aislamiento son diferentes? Supóngase un error de tipo I con α igual a 0.01.

Se desea probar la hipótesis nula de que

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu,$$

o en forma equivalente

$$H_0: \tau_j = 0, \qquad j = 1, 2, \ldots, 5.$$

Los tamaños de las muestras son $n_1 = 5$, $n_2 = 4$, $n_3 = 5$, $n_4 = 3$, y $n_5 = 6$; así que $N = 5 + 4 + \cdots + 6 = 23$. Las sumas de los tratamientos son $T_{\cdot 1} = 73.3$, $T_{\cdot 2} = 57.4$, $T_{\cdot 3} = 69.2$, $T_{\cdot 4} = 39.6$, y $T_{\cdot 5} = 78$. La suma total es $T = 73.3 + 57.4 + \cdots + 78 = 317.5$. Las sumas de los cuadrados son las siguientes:

$$\text{STC} = 14.4^2 + 14.8^2 + \cdots + 12.7^2 - \frac{317.5^2}{23} = 11.05,$$

$$\text{SCTR} = \frac{73.3^2}{5} + \frac{57.4^2}{4} + \frac{69.2^2}{5} + \frac{39.6^2}{3} + \frac{78^2}{6} - \frac{317.5^2}{23} = 9.836,$$

$$\text{SCE} = 11.05 - 9.836 = 1.214.$$

La información se ha agrupado en una tabla de análisis de varianza que se muestra en la tabla 12.4. Dado que $f = 36.48 > f_{0.99,\,4,\,18} = 4.58$ se rechaza la hipótesis nula de que no existe ningún efecto debido a los tratamientos. En relación con lo anterior, existe una razón para creer que parte de los consumos promedio de energía son diferentes para los cinco niveles de aislamiento.

TABLA 12.4 Tabla ANOVA para el ejemplo 12.1

Fuente de variación	*gl*	*SC*	*CM*	*Valor F*
Tratamientos	4	9.836	2.459	36.48
Error	18	1.214	0.0674	
Total	22	11.05	$f_{0.99,\,4,\,18} = 4.58$	

12.4.2 Método de Scheffé para comparaciones múltiples

Recuérdese que la hipótesis alternativa en el análisis de varianza no especifica qué medias son diferentes; lo que establece es que por lo menos una es diferente a las otras, así que el rechazo de la hipótesis nula con base en la estadística F no puede emplearse como fundamento para aceptar una alternativa en particular. Por ejemplo, supóngase que se rechaza la hipótesis nula H_0: $\mu_1 = \mu_2 = \mu_3$; lo anterior significa que μ_3 es diferente, pero que μ_1 y μ_2 son las mismas. O puede expresar que las tres medias son diferentes entre sí, o cualquier otra combinación posible de estos resultados. Por lo tanto, ésta es una razón muy fuerte para que el investigador necesite un análisis más completo para explorar las diferencias estadísticamente discernibles entre cierto número de medias de población.

Con este propósito se han propuesto varios métodos; entre éstos se encuentran el procedimiento de rangos estudentizados de Tukey, la prueba de rangos múltiples de Duncan y el métodos de Scheffé (véase [5]). Sólo se analizará el método de Scheffé para comparaciones múltiples debido a que tiene, en forma relativa, pocas restricciones y es preferido por muchos cuando se comparan combinaciones de las medias de los tratamientos. El método de Scheffé radica en la formulación de un *contraste* que es una comparación que escoge el investigador para representar una combinación lineal de cualquier número de medias de población. Un contraste es un método general de comparación que permite al investigador determinar, con base en la evidencia de la muestra, si el contraste dado es estadísticamente discernible.

Se define un contraste, denotado por L, como

$$L = \sum_{j=1}^{k} c_j \mu_j, \tag{12.11}$$

en donde μ_j es la media del j-ésimo nivel, y las c_j's son constantes tales que $\sum_{j=1}^{k} c_j = 0$. Por ejemplo, $L = \mu_3 - \mu_4$ es un contraste con $c_1 = 1$ y $c_2 = -1$. Este contraste es una comparación entre μ_3 y μ_4. Otro contraste es $L = 3\mu_1 - \mu_2 - \mu_3 - \mu_4$, con $c_1 = 3$, $c_2 = c_3 = c_4 = -1$. Este contraste es una comparación entre μ_1 y μ_2, μ_3, y μ_4. De esta forma el método de Scheffé permite que el investigador escoja las comparaciones basadas en las características de interés.

Un estimador no sesgado de L está dado por

$$\hat{L} = \sum_{j=1}^{k} c_j \bar{Y}_{\cdot j}, \tag{12.12}$$

cuya varianza se estima mediante

$$s^2(\hat{L}) = \text{CME} \sum_{j=1}^{k} \frac{c_j^2}{n_j}. \tag{12.13}$$

Scheffé demostró (véase [7]) que todos los posibles contrastes definidos por (12.11) se encuentran incluidos, con una probabilidad de $1 - \alpha$, en el conjunto de intervalos

$$\hat{L} - A s(\hat{L}) \leq L \leq \hat{L} + A s(\hat{L}), \tag{12.14}$$

en donde

$$A = \sqrt{(k - 1) f_{1-\alpha,\, k-1,\, N-k}}$$

y $\hat{L}$ y $s^2(\hat{L})$ se definen mediante (12.12) y (12.13), respectivamente. Si para algún contraste L se obtiene un intervalo a partir de (12.14) que no incluye al cero, entonces el contraste es estadísticamente discernible. Por lo tanto, en realidad para cada contraste L se está probando la hipótesis nula

$$H_0\colon L = 0.$$

La esencia del conjunto de intervalos definidos por (12.14) es que para *todos* los intervalos el nivel de confianza es de $100(1 - \alpha)$. Si se va a repetir un experimento muchas veces, y para cada una se calculan los invervalos de confianza para todos los posibles contrastes mediante el empleo de (12.14), entonces en un $100(1 - \alpha)$ de las repeticiones, todos los intervalos de confianza serán correctos. Que el intervalo de confianza sea del $100(1 - \alpha)$ para *todos* los intervalos, es mejor a obtener un intervalo de confianza del $100(1 - \alpha)$ para cada par de medias de tratamientos, en cuyo caso el nivel de confianza sólo es para cada par individual y no para el conjunto entero de éstos.

Ejemplo 12.2 En el ejemplo 12.1, compárese μ_4 contra μ_5; μ_2, μ_3, y μ_4 contra μ_5; μ_1 contra μ_2; y μ_3 y μ_4 contra μ_5, empleando el método de Scheffé con $\alpha = 0.01$.

Aunque pueden efectuarse comparaciones entre diversas combinaciones de los tratamientos, ciertas comparaciones parecen razonables si el objetivo es el ordenar los tratamientos en subgrupos dentro de los cuales no aparezca ninguna diferencia apreciable. Por ejemplo, si no existe una diferencia discernible entre el empleo de energía promedio para aislamientos térmicos de 10 y 12 pulgadas, puede ser, desde un punto de vista económico, más razonable utilizar un aislamiento de 10 pulgadas que uno de 12. Los contrastes para las cuatro comparaciones son:

$$L_1 = \mu_4 - \mu_5, \qquad L_2 = \mu_2 + \mu_3 + \mu_4 - 3\mu_5,$$

$$L_3 = \mu_1 - \mu_2, \qquad L_4 = 2\mu_5 - \mu_3 - \mu_4.$$

Se ilustrará el cálculo del intervalo de confianza para L_2. Dado que $\bar{y}_2 = 14.35$,

$\bar{y}_{\cdot 3} = 13.84$, $\bar{y}_{\cdot 4} = 13.2$, y $\bar{y}_{\cdot 5} = 13$,

$$\hat{L}_2 = 14.35 + 13.84 + 13.2 - (3)(13) = 2.39.$$

La varianza estimada es

$$s^2(\hat{L}_2) = 0.0674 \left[\frac{1^2}{4} + \frac{1^2}{5} + \frac{1^2}{3} + \frac{(-3)^2}{6}\right] = 0.1539,$$

y

$$s(\hat{L}_2) = 0.3923.$$

Dado que $f_{0.99, 4, 18} = 4.58$, $A = \sqrt{(4)(4.58)} = 4.28$, el intervalo de confianza para L_2 es

$$2.39 \pm (4.28)(0.3923) = (0.7109, 4.0691).$$

Al seguir el mismo procedimiento se obtiene que los intervalos de confianza para los otros contrastes son

$$L_1: \quad (-0.5857, 0.9857),$$

$$L_3: \quad (-0.4354, 1.0554),$$

$$L_4: \quad (-2.2572, 0.1772).$$

Nótese que de los cuatro intervalos de confianza para los contrastes de interés sólo el de L_2 no incluye el valor cero. Dado que la inclusión de este valor en estos intervalos de confianza es equivalente a la falta de significancia estadística en una prueba bilateral con respecto a la diferencia entre las medias, una comparación de los cuatro intervalos revela que no existe ninguna diferencia apreciable en el consumo de energía promedio para un grosor del aislamiento térmico de 8, 10 o 12 pulgadas. Se llega a esta conclusión debido a que los contrastes L_1 y L_4 no son estadísticamente discernibles, pero L_2 sí lo es. Dado que L_2 es igual que L_4 excepto que éste contiene a μ_2 (6 pulgadas de aislamiento), con base en los resultados de este experimento puede considerarse a un aislamiento de 8 pulgadas de espesor, como óptimo, desde un punto de vista económico.

Debe notarse que si se rechaza la hipótesis nula de medias iguales mediante el empleo de la estadística *F*, entonces el método de Scheffé dará por lo menos un contraste que es estadísticamente significativo.

12.4.3 Análisis de residuos y efectos de la violación de las suposiciones

De la sección 9.6.3. recuérdese que, para muestras de diferente tamaño, el efecto de violar la suposición de varianzas iguales cuando se comparan dos medias puede ser sustancial. Dado que esta misma suposición se formula cuando se comparan k medias, se desean examinar las formas en que lo anterior puede detectarse y analizar los efectos sobre la inferencia cuando no violan las suposiciones.

Una forma sencilla y útil para detectar la discrepancia con el modelo propuesto se basa en un análisis de residuos. Un *residuo* es un estimador del error aleatorio ε_{ij}. Dado que

$$\varepsilon_{ij} = Y_{ij} - \mu_j,$$

el residuo correspondiente denotado por e_{ij}, se define como

$$e_{ij} = y_{ij} - \bar{y}_{\cdot j}, \qquad j = 1, 2, \ldots, k, \qquad i = 1, 2, \ldots, n_j.$$

Los residuos no son estimados en el sentido de estimación de parámetros, sino como estimadores de los valores de las variables aleatorias no observables ε_{ij} con base en los estimadores $\bar{y}_{\cdot j}$ para los k medias de población.

Si es válida la suposición de que los errores aleatorios tienen las mismas varianzas para todos los niveles de k, entonces una gráfica de los residuos de cada tratamiento no revelará ninguna diferencia apreciable en la dispersión de los residuos alrededor del cero. Si esta dispersión es notablemente diferente para algunos tratamientos, entonces es posible que las varianzas no sean iguales para todos los tratamientos. Para normalizar la escala de magnitudes de los residuos es preferible emplear los *residuos estandarizados* $e_{ij}/\sqrt{\text{CME}}$. Entonces, dado que por hipótesis los errores aleatorios se encuentran normalmente distribuidos, un residuo estandarizado rara vez se encontrará más allá de un intervalo de ± 3

Se ilustrará el análisis de residuos empleando los datos del ejemplo 12.1. Dado que $\bar{y}_{\cdot 1} = 14.66$ y $\sqrt{\text{CME}} = 0.2596$, los residuos para el primer tratamiento son $14.4 - 14.66 = -0.26$, $14.8 - 14.66 = 0.14$, $15.2 - 14.66 = 0.54$, $14.3 - 14.66 = -0.36$, y $14.6. - 14.66 = -0.06$, y los residuos correspondientes estandarizados son –1.00, 0.54, 2.08, –1.39 y –0.23. Al seguir este procedimiento se obtienen todos los residuos estandarizados que aparecen en la tabla 12.5.

La figura 12.1 ilustra los residuos estandarizados para cada tratamiento. Se observa que no existe ninguna diferencia notable en la dispersión para cada uno de los cinco tratamientos excepto para uno de los residuos del primer tratamiento. De acuerdo con lo anterior, parece que la hipótesis de que las varianzas de los cinco tratamientos son las mismas, es razonable en este caso. También se encuentran disponibles en la literatura estadística procedimientos formales para verificar la hipótesis de igualdad entre las k varianzas. Dos de los usados con más frecuencia son la prueba de Bartlett y la prueba de Hartley. Se invita al lector a que consulte [5] para conocer los detalles.

TABLA 12.5 Residuos estandarizados para el ejemplo 12.1

4	6	8	10	12
−1.00	0.58	−0.15	−0.77	0.39
0.54	−0.96	1.00	0.77	−0.77
2.08	0.96	−0.54	0	−0.39
−1.39	−0.58	−0.92		0.77
−0.23		0.62		1.16
				−1.16

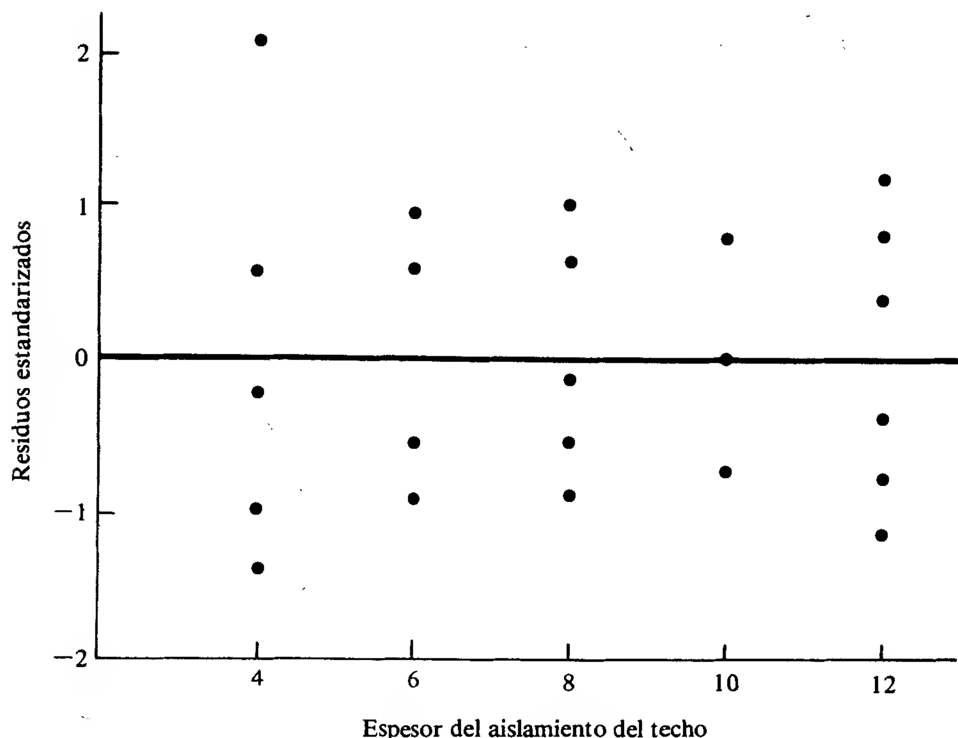


FIGURA 12.1 Gráfica de los residuos estandarizados para los cinco tratamientos del ejemplo 12.1

Como se examinó en el capítulo 9, el efecto sobre las inferencias con respecto a las medias, cuando los errores aleatorios no se encuentran normalmente distribuidos, es menor mientras el alejamiento de la normalidad no sea muy severo. De esta forma, la estadística F en el análisis de varianza es robusta con respecto a los alejamientos de la hipótesis de normalidad. Si las varianzas de todos los tratamientos no son iguales entre sí, puede aumentarse el tamaño de la región crítica de la estadística F para el caso de efectos fijos; pero, como se analizó en el capítulo 9, este efecto puede minimizarse mediante el empleo de muestras de igual tamaño para cada tratamiento. En otras palabras, en el análisis de varianza, la estadística F también es más robusta ante varianzas desiguales siempre y cuando los tamaños de la muestra de los tratamientos sean iguales. Desafortunadamente este resultado no se extiende al caso de efectos aleatorios en el que la violación de la hipótesis de varianzas iguales generalmente tendrá efectos considerables sobre las inferencias aun para muestras del mismo tamaño.

La hipótesis crucial en el desarrollo del análisis de varianza es que los errores aleatorios son independientes. Si los errores son interdependientes, el tamaño real de la región crítica puede ser, en forma substancial, más grande (cinco o más veces) que

el tamaño dictado al seleccionar la probabilidad del error de tipo I. Se invita al lector a que consulte [3], para una revisión de las consecuencias que surgen al violar las suposiciones en el análisis de varianza.

12.4.4 El caso de efectos aleatorios

Para introducir el caso de efectos aleatorios se utilizará el siguiente análisis breve. Para una presentación más completa se sugiere consultar [6]. Para el modelo de efectos aleatorios se formuló la suposición de que los niveles empleados en el experimento fueron seleccionados en forma aleatoria de una población de posibles niveles. Además se supondrá que $\tau_j \sim N(0,\sigma_\tau^2)$, en donde σ_τ^2 es la varianza de los tratamientos aleatorios τ_j. La descomposición de la suma total de cuadrados y el análisis de varianza es igual a la del caso de efectos fijos para un experimento con sólo un factor, pero en este caso el valor esperado del cuadrado medio de tratamiento es diferente. Dadas muestras de igual tamaño n para todos los niveles, se puede demostrar que

$$E(\text{CME}) = \sigma^2,$$

y (12.15)

$$E(\text{CMTR}) = \sigma^2 + n\sigma_\tau^2.$$

La región apropiada de rechazo sigue siendo la misma ya que un valor grande del cociente entre *CMTR* y *CME* sugiere que debe rechazarse la hipótesis nula H_0: $\sigma_\tau^2 = 0$

Ejemplo 12.3 Una planta de enlatado emplea un número muy grande de máquinas para su proceso de llenado. Se da por hecho que cada máquina vacía un peso especificado del producto en cada lata. El gerente de la planta sospecha que existe una gran variación en la cantidad del producto que se vacía entre las distintas máquinas. Para verificar su sospecha, escoge al azar cuatro máquinas y pesa el contenido de cinco latas, seleccionadas en forma aleatoria, llenadas por cada una de las cuatro máquinas. Los resultados se muestran en la tabla 12.6. ¿Qué proporción de la varianza en los pesos puede atribuirse a las diferencias que existen entre las máquinas?

Primero se llevará a cabo un análisis de varianza para saber si puede rechazarse H_0: $\sigma_\tau^2 = 0$. Los totales de las máquinas son $T_{.1} = 6.14$, $T_{.2} = 6.03$, $T_{.3} = 5.99$ y

TABLA 12.6 Contenido en peso para un proceso de llenado

Máquina			
1	*2*	*3*	*4*
1.24	1.20	1.19	1.18
1.22	1.20	1.20	1.18
1.22	1.21	1.19	1.19
1.23	1.22	1.20	1.18
1.23	1.20	1.21	1.20

TABLA 12.7 Tabla ANOVA para el ejemplo 12.3

Fuente de variación	gl	SC	CM	*Valor F*
Tratamientos	3	0.004695	0.001565	20.87
Error	16	0.0012	0.000075	
Total	19	0.005895	$f_{0.95,\,3,\,16} = 3.24$	

$T_{\cdot 4} = 5.93$. El total global es $T_{\cdot\cdot} = 24.09$, y los tamaños de todas las muestras son $n = 5$. Entonces

$$STC = 1.24^2 + 1.22^2 + \cdots + 1.20^2 - \frac{24.09^2}{20} = 0.005895,$$

$$SCTR = \frac{6.14^2 + 6.03^2 + 5.99^2 + 5.93^2}{5} - \frac{24.09^2}{20} = 0.004695,$$

$$SCE = 0.005895 - 0.004695 = 0.0012.$$

La tabla ANOVA se da en la tabla 12.7. Dado que $f = 20.87 > f_{0.95,\,3,\,16} = 3.24$, se rechaza la hipótesis nula de que no hay variación debida a las máquinas.

Para estimar la varianza en los pesos y qué proporción de ésta puede atribuirse a las diferencias entre las máquinas, recuérdese que para un modelo de efectos aleatorios

$$Var(Y_{ij}) = \sigma^2 + \sigma_\tau^2.$$

De (12.15), un estimado de σ^2 es $CME = 0.000075$, y un estimador de $\sigma^2 + 5\sigma_\tau^2$ es $CMTR = 0.001565$. En otras palabras,

$$0.000075 + 5s_\tau^2 = 0.001565$$

$$s_\tau^2 = \frac{0.001565 - 0.000075}{5}$$

$$= 0.000298$$

es un estimador de σ_τ^2. Entonces un estimador de la varianza en el peso es

$$s^2(Y_{ij}) = 0.000075 + 0.000298$$

$$= 0.000373,$$

de la cual 0.000298/0.000373, o el 79.89%, se debe a diferencias entre las máquinas.

12.5 Análisis de experimentos con sólo un factor en un diseño en bloque completamente aleatorizado

Recuérdese que cuando las unidades experimentales no son homogéneas, se introduce una fuente potencial de variación que, en general, puede afectar la inferencia con respecto al factor de interés. En estos casos es necesario emplear un diseño aleatorizado para remover la fuente externa de variación con lo que se incrementa la sensitividad para detectar diferencias entre los tratamientos de interés.

Ejemplo 12.4 La agencia de Protección del Medio Ambiente (APMA) anualmente clasifica de acuerdo con la eficiencia en el quemado de combustible a todos los automóviles disponibles para venta de Estados Unidos. Sin embargo, es un hecho muy conocido que las clasificaciones de la APMA se basan, principalmente, en pruebas de laboratorio y de esta forma se tiende a sobreestimar la eficiencia real en el quemado de combustible. Una empresa independiente desea determinar si existe una diferencia, estadísticamente discernible, en la eficiencia del quemado promedio de combustible bajo condiciones de rodamiento real para cinco automóviles compactos que tienen la misma clasificación APMA. La empresa tiene acceso a un recorrido de 400 millas que incluye tanto el manejo en ciudad como en carretera. Estúdiense los aspectos de diseño de este experimento.

Es claro que los tratamientos están constituidos por los cinco automóviles y que la respuesta medible es el número de millas por galón logradas por los automóviles durante el recorrido de 400 millas. Pero, ¿cuál es la unidad experimental?; ésta tiene que ser la persona que maneja el automóvil, pero no es común que una empresa que realiza pruebas utilice un conductor para todo el experimento. Supóngase que se escogen cuatro conductores para el experimento. Aunque la empresa explicará el propósito del experimento en forma breve, a los conductores ya se ha introducido otra fuente de posible variación. No importa qué tan similares sean los conductores entre sí; a pesar de todo existe un riesgo potencial de tener efectos por los conductores que pueden tomarse en cuenta mediante la creación de cuatro bloques, uno para cada conductor, de tal manera que los tratamientos dentro de cada bloque (los cinco automóviles) se apliquen a unidades experimentales homogéneas (el mismo conductor). La pregunta que surge en este momento es, ¿cómo asignar los automóviles a los conductores? El diseño aleatorizado especifica que la asignación de los tratamientos a las unidades experimentales dentro de cada bloque debe hacerse en forma aleatoria. De esta manera, para asignar el orden en el cual serán manejados los automóviles por cada conductor, se concibe un proceso de selección aleatorio simple. Por ejemplo, la asignación puede hacerse de acuerdo con la tabla 12.8, la cual constituye un diseño en bloque completamente aleatorizado.

A continuación se analizará un experimento con sólo un factor en un diseño en bloques completamente aleatorizado. Primero será necesario generalizar para después regresar al ejemplo de la eficiencia en consumo de combustible e ilustrar los pasos del cálculo. Las observaciones del experimento pueden colocarse como se muestran en la tabla 12.9.

TABLA 12.8 Diseño en bloque completamente aleatorizado para el ejemplo 12.4

		Automóvil				
Conductor	1	A_1	A_3	A_5	A_4	A_2
	2	A_5	A_3	A_4	A_2	A_1
	3	A_4	A_1	A_5	A_3	A_2
	4	A_2	A_5	A_4	A_1	A_3

Supóngase que se tienen k tratamientos y n bloques, el modelo matemático para un diseño con sólo un factor en bloques completamente aleatorizado es

$$Y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij} \qquad i = 1, 2, \ldots, n, \qquad (12.16)$$
$$j = 1, 2, \ldots, k,$$

en donde Y_{ij} es la observación de la respuesta en el i-ésimo bloque y bajo el j-ésimo tratamiento, μ es la media global, β_i es el efecto sobre la respuesta debido al i-ésimo bloque, τ_j es el efecto debido al j-ésimo tratamiento y ε_{ij} es el error aleatorio. Como en el caso anterior, se da por hecho que los errores son variables aleatorias independientes, tales que $\varepsilon_{ij} \sim N(0, \sigma^2)$ para toda i y j. Si tanto los tratamientos como los bloques son de efectos fijos, entonces las β_i'y los τ_j' son parámetros fijos que representan desviaciones de las medias de los bloques y los tratamientos de la media global, respectivamente. En otras palabras,

$$\beta_i = \mu_{i\cdot} - \mu, \qquad i = 1, 2, \ldots, n, \qquad (12.17)$$
$$\tau_j = \mu_{\cdot j} - \mu, \qquad j = 1, 2, \ldots, k,$$

en donde $\mu_{i\cdot}$ y $\mu_{\cdot j}$ son las medias de las población para el i-ésimo bloque y el j-ésimo tratamiento, respectivamente.

Al igual que en el diseño completamente aleatorizado, se supone que las varianzas de la población para todos los tratamientos son iguales. También debe suponerse que el efecto del tratamiento sobre la respuesta es el mismo para todos los bloques; en otras

TABLA 12.9 Arreglo común de las observaciones para un diseño con sólo un factor en bloque completamente aleatorizado

		Tratamiento					
		1	2	⋯	j	⋯	k
Bloque	1	Y_{11}	Y_{12}	⋯	Y_{1j}	⋯	Y_{1k}
	2	Y_{21}	Y_{22}	⋯	Y_{2j}	⋯	Y_{2k}
	⋮	⋮	⋮		⋮		⋮
	i	Y_{i1}	Y_{i2}	⋯	Y_{ij}	⋯	Y_{ik}
	⋮	⋮	⋮		⋮		⋮
	n	Y_{n1}	Y_{n2}	⋯	Y_{nj}	⋯	Y_{nk}

palabras, puede obtenerse la misma conclusión a partir de todos los bloques con respecto al efecto del tratamiento. Cuando esto ocurre se dice que los tratamientos y los bloques *no interactúan*, y sus efectos individuales sobre la respuesta son aditivos. La noción de *interacción* entre dos factores se examinará en la siguiente sección.

Para un diseño de un sólo factor en bloque completamente aleatorizado, el principal propósito es determinar si las diferencias en los tratamientos son estadísticamente significativas, es decir, para el caso de efectos fijos se desea probar la hipótesis nula

$$H_0: \tau_j = 0, \qquad j = 1, 2, \ldots, k.$$

El lector puede sorprenderse con respecto al efecto del bloque, pero el interés, en realidad, no recae en determinar si éste es estadísticamente apreciable. Todo lo que se desea hacer es aislar el efecto del bloque y removerlo del error experimental, de tal manera que se incremente la eficiencia para detectar diferencias reales entre los tratamientos, si es que éstas existen.

Para el procedimiento del análisis de varianza, puede escribirse el modelo dado por (12.16) como

$$\varepsilon_{ij} = Y_{ij} - \mu - \beta_i - \tau_j. \tag{12.18}$$

Al sustituir (12.17) para β_i y τ_j en (12.18), se tiene

$$\varepsilon_{ij} = Y_{ij} - \mu - \mu_{i\cdot} + \mu - \mu_{\cdot j} + \mu. \tag{12.19}$$

Ahora, al reemplazar (12.17) para β_i y τ_j y (12.19) para ε_{ij} en (12.16) se obtiene la siguiente identidad:

$$Y_{ij} - \mu = (\mu_{i\cdot} - \mu) + (\mu_{\cdot j} - \mu) + (Y_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu). \tag{12.20}$$

En otras palabras, la desviación de una observación con respecto a la media global tiene tres componentes (la desviación debida a los bloques, a los tratamientos y al error aleatorio).

Para las observaciones que se encuentran en la tabla 12.9 se definen las siguientes estadísticas:

$$T_{i\cdot} = \sum_{j=1}^{k} Y_{ij}, \qquad \overline{Y}_{i\cdot} = T_{i\cdot}/k, \qquad i = 1, 2, \ldots, n$$

$$T_{\cdot j} = \sum_{i=1}^{n} Y_{ij}, \qquad \overline{Y}_{\cdot j} = T_{\cdot j}/n, \qquad j = 1, 2, \ldots, k$$

$$T_{\cdot\cdot} = \sum_{i=1}^{n} \sum_{j=1}^{k} Y_{ij}, \qquad \overline{Y}_{\cdot\cdot} = T_{\cdot\cdot}/nk.$$

Por lo tanto, la identidad en términos de la muestra correspondiente a (12.20) es

$$Y_{ij} - \overline{Y}_{\cdot\cdot} = (\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}) + (\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot}) + (Y_{ij} - \overline{Y}_{i\cdot} - \overline{Y}_{\cdot j} + \overline{Y}_{\cdot\cdot}).$$

Al elevar al cuadrado ambos miembros y llevar a cabo la suma sobre i y j se tiene la relación

$$\sum_{i=1}^{n}\sum_{j=1}^{k}(Y_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^{n}\sum_{j=1}^{k}(\bar{Y}_{i.} - \bar{Y}_{..})^2 + \sum_{i=1}^{n}\sum_{j=1}^{k}(\bar{Y}_{.j} - \bar{Y}_{..})^2 + \sum_{i=1}^{n}\sum_{j=1}^{k}(Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2,$$

en donde puede demostrarse que los tres términos que contienen productos cruzados se reducen a cero. Ésta es la ecuación fundamental para el análisis de varianza, y establece que la suma total de los cuadrados *STC* se separa en la suma de los cuadrados de los bloques *SCB*, la suma de los cuadrados de los tratamientos SCTR y la suma de los cuadrados de los errores SCE.

Por causa de la restricción $\sum_{i=1}^{n}\sum_{j=1}^{k}(Y_{ij} - \bar{Y}_{..}) = 0$, el número de grados de libertad para STC es igual a $nk - 1$. En forma similar, por causa de las restricciones $\sum_{i=1}^{n}(\bar{Y}_{i.} - \bar{Y}_{..}) = 0$ y $\sum_{j=1}^{k}(\bar{Y}_{.j} - \bar{Y}_{..}) = 0$, el número de grados de libertad para *SCB* y *SCTR* son iguales a $n - 1$ y $k - 1$, respectivamente. Se sigue que

$$\begin{aligned} \text{gl(SCE)} &= \text{gl(STC)} - \text{gl(SCB)} - \text{gl(SCTR)} \\ &= nk - 1 - (n - 1) - (k - 1) \\ &= (n - 1)(k - 1). \end{aligned}$$

Puede demostrarse que bajo las suposiciones del modelo y la hipótesis H_0: $\tau_j = 0$, SCTR/σ^2 y SCE/σ^2 son dos variables aleatorias independientes con una distribución chi-cuadrada con $k - 1$ y $(n - 1)(k - 1)$ grados de libertad, en forma correspondiente. También puede demostrarse que los valores esperados de los cuadrados medios del error y del tratamiento son

$$E(\text{CME}) = \sigma^2$$

y

$$E(\text{CMTR}) = \sigma^2 + \frac{n\sum_{j=1}^{k}\tau_j^2}{k - 1}.$$

Entonces, con base en el argumento previo, la estadística de prueba apropiada es el cociente de los cuadrados medios del tratamiento y del error, el cual tiene una distribución F con $k - 1$ y $(n - 1)(k - 1)$ grados de libertad. Como antes, se sugiere una región crítica de tamaño α, ya que un valor grande del cociente tiende a implicar que no todas las medias de los tratamientos son las mismas. El análisis de varianza aparece en la tabla 12.10.

Debe notarse que es posible una prueba para el efecto de bloque al formar el cociente entre *CMB* y *CME* y compararlo con la región crítica que se encuentra en el extremo superior de una distribución F con $n - 1$ y $(n - 1)(k - 1)$ grados de libertad.

TABLA 12.10 Tabla de análisis de varianza para un experimento con sólo un factor en bloque completamente aleatorizado

Fuente de variación	gl	SC	CM	*Estadística F*
Bloques	$n - 1$	$\Sigma\Sigma(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$		
Tratamientos	$k - 1$	$\Sigma\Sigma(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2$	CMTR = SCTR/$(k - 1)$	$F = \dfrac{\text{CMTR}}{\text{CME}}$
Error	$(n - 1)(k - 1)$	$\Sigma\Sigma(Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot})^2$	CME = SCE/$(n - 1)(k - 1)$	
Total	$nk - 1$	$\Sigma\Sigma(Y_{ij} - \bar{Y}_{\cdot\cdot})^2$		

Lo anterior no constituye en realidad una parte integral del análisis. Después de todo, se escoge un bloque completamente aleatorizado para un experimento con sólo un factor para remover el efecto potencial de la fuente de variación extraña. Si tal efecto es estadísticamente significativo, realmente no es de gran interés.

Para realizar cálculos a mano, es preferible emplear las siguientes fórmulas que son equivalentes, en un sentido algebraico, para obtener las sumas de cuadrados.

$$\text{STC} = \sum_{i=1}^{n}\sum_{j=1}^{k} y_{ij}^2 - \frac{T_{\cdot\cdot}^2}{nk}$$

$$\text{SCB} = \frac{1}{k}\sum_{i=1}^{n} T_{i\cdot}^2 - \frac{T_{\cdot\cdot}^2}{nk}$$

$$\text{SCTR} = \frac{1}{n}\sum_{j=1}^{k} T_{\cdot j}^2 - \frac{T_{\cdot\cdot}^2}{nk}$$

$$\text{SCE} = \text{STC} - \text{SCB} - \text{SCTR}$$

Para ilustrar los pasos de cálculo, supóngase que los resultados del experimento descrito en el ejemplo 12.4 son los que se muestran en la tabla 12.11 (las mediciones están dadas en millas por galón para un recorrido de 400 millas). Para probar la hipótesis nula

$$H_0\colon \tau_j = 0, \quad j = 1, 2, \ldots, 5,$$

las sumas de cuadrados dan

$$\text{STC} = 33.6^2 + 36.9^2 + \cdots + 32.8^2 - \frac{672.4^2}{20} = 102.212,$$

$$\text{SCB} = \frac{156.1^2 + \cdots + 172.4^2}{5} - \frac{672.4^2}{20} = 41.676,$$

$$\text{SCTR} = \frac{139.5^2 + \cdots + 133.3^2}{4} - \frac{672.4^2}{20} = 38.092,$$

TABLA 12.11 Datos experimentales para el ejemplo 12.4

Conductor	*Automóvil* A_1	A_2	A_3	A_4	A_5	*Totales*
1	33.6	32.8	31.9	27.2	30.6	$T_{1.} = 156.1$
2	36.9	36.1	32.1	34.4	35.3	$T_{2.} = 174.8$
3	34.2	35.3	33.7	31.3	34.6	$T_{3.} = 169.1$
4	34.8	37.1	34.8	32.9	32.8	$T_{4.} = 172.4$
Totales	$T_{.1} = 139.5$	$T_{.2} = 141.3$	$T_{.3} = 132.5$	$T_{.4} = 125.8$	$T_{.5} = 133.3$	$T_{..} = 672.4$

$$\text{SCE} = 102.212 - 41.676 - 38.092 = 22.444.$$

La tabla ANOVA se encuentra dada en la tabla 12.12. Dado que $f = 5.09 > f_{0.95,\ 4,\ 12} = 3.26$, se rechaza la hipótesis nula de igualdad de efecto de tratamiento. Por lo tanto, existe una razón para creer que las eficiencias en consumo medio de combustible de algunos de estos automóviles no son iguales.

La identificación y eliminación del efecto de los bloques de la variación total permite que se hagan comparaciones múltiples sobre los tratamientos, como ya se vio en la sección 12.4.2. Pueden definirse y probarse un gran número de contrastes para determinar si son estadísticamente apreciables al seguir el procedimiento delineado en la sección 12.4.2. La única excepción es que la cantidad denotada por A en (12.14) ahora está dada por

$$A = \sqrt{(k - 1)f_{1-\alpha,\ k-1,\ (n-1)(k-1)}}.$$

A veces los bloques no son de efectos fijos, es decir, se eligen para el experimento en forma aleatoria de una población de posibles bloques. Si los tratamientos son de efectos fijos, la única diferencia con respecto al caso previo se encuentra en la suposición de β_i; i.e., $\beta_i \sim N(0,\ \sigma_\beta^2)$; pero el análisis sigue siendo el mismo, aun para comparaciones múltiples entre los tratamientos.

Además de la suposición de independencia, se hacen dos suposiciones clave para un diseño en bloques aleatorizados: las varianzas de cada tratamiento son iguales y

TABLA 12.12 Tabla ANOVA para el ejemplo 12.4

Fuente de variación	gl	SC	CM	*Valor F*
Bloques	3	41.676		
Tratamientos	4	38.092	9.523	5.09
Error	12	22.444	1.870	
Total	19	102.212	$f_{0.95,\ 4,\ 12} = 3.26$	

los bloques y tratamientos no interactúan. La presencia de interacción entre bloques y tratamientos implica que no es posible evaluar el efecto del tratamiento sobre todos los bloques, sino que éste se debe describir en forma individual para cada bloque. Si además los efectos del bloque y del tratamiento son aditivos, la estadística F no es sensitiva a la violación de la suposición de varianzas iguales; para éstas, si existe una interacción entre bloques y tratamientos, la estadística F se encuentra sesgada negativamente, es decir, si se rechaza la hipótesis nula de que no existe diferencia alguna entre los tratamientos, entonces puede confiarse en que existe una diferencia entre los tratamientos. Pero si la hipótesis nula no se rechaza, esto se puede deber, ya sea a un sesgo negativo (la presencia de interacción) o a la ausencia de diferencias entre los tratamientos. Puede emplearse un procedimiento desarrollado por Tukey, el cual se describe en [4], para probar la interacción entre bloques y tratamientos.

Si se violan tanto la suposición de varianzas iguales como la de aditividad, la estadística F para las diferencias en los tratamientos tiene un sesgo positivo; en otras palabras, si se rechaza la hipótesis nula de que no existe ninguna diferencia entre los tratamientos, esto no necesariamente implica que las diferencias entre los tratamientos sean estadísticamente significativas. Cuando existe preocupación sobre estas suposiciones, debe usarse una *prueba F conservadora* desarrollada por Geisser y Greenhouse (véase [4]). Los pasos de cálculo para esta prueba son iguales a los del método convencional ya descrito, excepto que el número de grados de libertad para este caso es de 1 y $n - 1$ en lugar de $k - 1$ y $(n - 1)(k - 1)$, para cada uno. Si para ambas pruebas se rechaza la hipótesis nula, puede tenerse la seguridad de que las diferencias entre los tratamientos son estadísticamente significativas. Si ambas pruebas no rechazan a H_0, entonces se puede proceder como si no existiese diferencia alguna entre los tratamientos.

12.6 Experimentos factoriales

Hasta este momento la presentación se ha dirigido hacia el análisis del efecto de un factor sobre la variable respuesta. Pero en muchas situaciones prácticas es necesario investigar, en forma simultánea, los efectos que tienen varios factores sobre la respuesta. Una forma muy eficiente de lograr lo anterior es mediante el uso de un experimento factorial en el que todos los niveles de un factor se combinan con todos los niveles de cualquier otro para formar los tratamientos. Por ejemplo, en un experimento factorial de dos factores en el que uno tiene tres niveles y el otro dos, existirán $3 \times 2 = 6$ tratamientos. En otras palabras, la respuesta será observada bajo seis tratamientos diferentes.

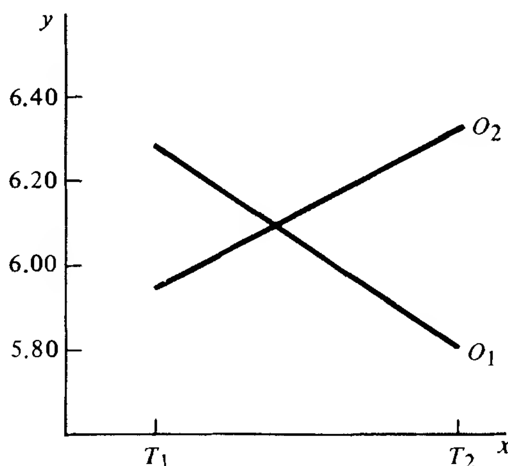
Con los experimentos factoriales no sólo es posible evaluar los efectos individuales de los factores sobre la respuesta, sino que también es posible determinar el efecto causado por sus interacciones. El efecto de un factor sobre una respuesta es simplemente el cambio en ésta, causado por un cambio en el nivel del factor. Pero si el efecto de un factor sobre la respuesta es diferente para distintos niveles de otro factor, entonces se dice que los dos factores interactúan entre sí. La presencia de interacción indica que el efecto de los factores sobre la respuesta es no lineal y de esta forma no puede asumirse un modelo aditivo.

Para ilustrar la interacción entre dos factores, considérese lo siguiente. Un fabricante de partes electrónicas emplea dos hornos y dos temperaturas con el propósito de probar la duración de cierto componente. Se seleccionan cuatro componentes de algún lote y se prueba su duración de acuerdo con las cuatro combinaciones posibles de hornos y temperaturas. El tiempo de duración de los componentes en horas es el siguiente:

	O_1	O_2
T_1	6.29	5.95
T_2	5.80	6.32

Los tratamientos para las cuatro posibles combinaciones de hornos y temperaturas son: O_1T_1, O_1T_2, O_2T_1, y O_2T_2. La diferencia en duración para los tratamientos O_1T_2 y O_1T_1 representa un estimador del efecto en la duración de los componentes en el primer horno, a consecuencia de un cambio en la temperatura. Se observa que este estimador es $5.80 - 6.29 = -0.49$. La diferencia en duración para los tratamientos O_2T_2 y O_2T_1 también es un estimador del efecto de la temperatura sobre la duración, pero ahora en el segundo horno. Esta diferencia es de $6.32 - 5.95 = 0.37$. Dado que estos dos estimadores son bastantes diferentes entre sí, el efecto de la temperatura en la duración del componente depende del horno en que éste se coloque. De esta forma, existe una interacción entre el horno y la temperatura. También se observa la misma ocurrencia al estimar el efecto del horno para $T_1(5.95 - 6.29 = -0.34)$ y $T_2(6.32 - 5.80 = 0.52)$. Estos resultados se ilustran en forma gráfica en la figura 12.2 en donde el eje y representa las observaciones de la respuesta; el eje x repre-

FIGURA 12.2 Efectos que interactúan



senta los niveles de un factor y los puntos graficados representan a cada nivel del otro factor. Si existe poca interacción entre el horno y la temperatura, las líneas que aparecen en la gráfica serían casi paralelas.

La determinación de si los efectos individuales o interacciones son estadísticamente apreciables puede hacerse sólo mediante inferencia estadística y no mediante el empleo de un analisis gráfico. En los siguientes párrafos se examinará un modelo no aditivo para un experimento factorial de dos factores en un diseño completamente aleatorizado. Se pueden analizar experimentos factoriales con más de dos factores mediante la extensión del procedimiento que a continuación se examina.

En un experimento factorial que incluye dos factores A y B con a y b niveles, respectivamente, el número de tratamientos es igual a $a \times b$. Si no se puede suponer un modelo aditivo (no interacción), sólo es posible una prueba para determinar si un efecto por interacción es estadísticamente apreciable, si se toma más de una observación de la respuesta para cada tratamiento. Lo anterior se debe a que no puede determinarse para cada estimador de la variación del error aleatorio a menos que la respuesta se observe más de una vez cada tratamiento, es decir, la evaluación de la variación del error aleatorio se basa en las diferencias en la respuesta observada bajo el mismo tratamiento. No está por demás notar que para un diseño completamente aleatorizado, los tratamientos deben aplicarse a unidades experimentales homogéneas sin importar cuántas veces se repita el proceso.

Si se suponen n aplicaciones de los ab tratamientos, el modelo matemático no aditivo para un factorial de dos factores es

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \qquad \begin{aligned} i &= 1, 2, \dots, a, \\ j &= 1, 2, \dots, b, \\ k &= 1, 2, \dots, n, \end{aligned} \tag{12.21}$$

en donde Y_{ijk} es la k-ésima observación de la respuesta para el tratamiento (i, j), μ es la media global, α_i es el efecto principal causado por el i-ésimo nivel de A, β_j es el efecto principal causado por el j-ésimo nivel de B, $(\alpha\beta)_{ij}$ es el efecto de interacción para el i-ésimo nivel de A y el j-ésimo nivel de B y ε_{ijk} es el k-ésimo error aleatorio en el tratamiento (i, j). Como antes, se supone que las varianzas de la población para cada uno de los ab tratamientos son iguales, y que los errores aleatorios son variables aleatorias independientes, normalmente distribuidas, con medias iguales a cero y varianza común σ^2.

Si se supone que los factores A y B son de efectos fijos, entonces α_i, β_j, y $(\alpha\beta)_{ij}$ son parámetros fijos, tales que

$$\sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = 0$$

y

$$\sum_{i=1}^{a} (\alpha\beta)_{ij} = \sum_{j=1}^{b} (\alpha\beta)_{ij} = 0.$$

para toda

Las siguientes hipótesis son de interés:

1. $H_0: (\alpha\beta)_{ij} = 0$ para toda i y j,
2. $H_0: \alpha_i = 0$ para toda i,
3. $H_0: \beta_j = 0$ para toda j.

Las últimas dos hipótesis incluyen los efectos (individuales) *principales* de los factores A y B, y la primera hipótesis pertenece a la posible interacción entre A y B. Si existe una fuerte interacción entre A y B, los resultados de las pruebas para demostrar un efecto principal causado por A o B pueden no ser significativos. Lo anterior es cierto debido a que los dos factores pueden interaccionar en tal forma (direcciones opuestas) que los efectos se compensen para uno o ambos factores. Este proceso de compensación puede evitar la detección de efectos principales significativos con base en una comparación entre las medias del nivel del factor.

Para desarrollar el procedimiento del análisis de varianza, puede escribirse el modelo (12.21) en términos de las desviaciones, al igual que en los casos previos.

$$Y_{ijk} - \mu = (\mu_{i\cdot\cdot} - \mu) + (\mu_{\cdot j\cdot} - \mu) + (\mu_{ij\cdot} - \mu_{i\cdot\cdot} - \mu_{\cdot j\cdot} + \mu) + (Y_{ijk} - \mu_{ij\cdot}), \quad (12.22)$$

en donde $\mu_{i\cdot\cdot}$ es la media real del i-ésimo nivel de A, $\mu_{\cdot j\cdot}$ es la media real del j-ésimo nivel de B y $\mu_{ij\cdot}$ es la media real del tratamiento (i, j). De esta forma, la igualdad dada por (12.22) establece que la desviación de una observación con respecto al promedio global está formada por cuatro componentes: las desviaciones causadas por el efecto principal de A; por el efecto principal de B; por el efecto de interacción entre A y B, por el error aleatorio.

Las observaciones de un factorial con dos factores en un experimento completamente aleatorizado pueden colocarse como se muestra en la tabla 12.13. De ésta se

TABLA 12.13 Arreglo común de las observaciones para un diseño factorial con dos factores y n observaciones por tratamiento

		A Nivel 1		A Nivel i		A Nivel a
	Nivel 1	$Y_{111} \cdots Y_{11k} \cdots Y_{11n}$	$\cdots$	$Y_{i11} \cdots Y_{i1k} \cdots Y_{i1n}$	$\cdots$	$Y_{a11} \cdots Y_{a1k} \cdots Y_{a1n}$
	⋮	⋮		⋮		⋮
B	Nivel j	$Y_{1j1} \cdots Y_{1jk} \cdots Y_{1jn}$	$\cdots$	$Y_{ij1} \cdots Y_{ijk} \cdots Y_{ijn}$	$\cdots$	$Y_{aj1} \cdots Y_{ajk} \cdots Y_{ajn}$
	⋮	⋮		⋮		⋮
	Nivel b	$Y_{1b1} \cdots Y_{1bk} \cdots Y_{1bn}$	$\cdots$	$Y_{ib1} \cdots Y_{ibk} \cdots Y_{ibn}$	$\cdots$	$Y_{ab1} \cdots Y_{abk} \cdots Y_{abn}$

definen las siguientes estadísticas:

$$T_{i\cdot\cdot} = \sum_{j=1}^{b} \sum_{k=1}^{n} Y_{ijk}, \qquad T_{\cdot j\cdot} = \sum_{i=1}^{a} \sum_{k=1}^{n} Y_{ijk}, \qquad T_{\cdot\cdot k} = \sum_{i=1}^{a} \sum_{j=1}^{b} Y_{ijk},$$

$$\overline{Y}_{i\cdot\cdot} = T_{i\cdot\cdot}/nb, \qquad \overline{Y}_{\cdot j\cdot} = T_{\cdot j\cdot}/na, \qquad \overline{Y}_{\cdot\cdot k} = T_{\cdot\cdot k}/ab,$$

$$T_{ij\cdot} = \sum_{k=1}^{n} Y_{ijk}, \qquad \overline{Y}_{ij\cdot} = T_{ij\cdot}/n,$$

$$T_{\cdot\cdot\cdot} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} Y_{ijk}, \qquad \overline{Y}_{\cdot\cdot\cdot} = T_{\cdot\cdot\cdot}/nab.$$

Nótese que $T_{i\cdot\cdot}$ ($T_{\cdot j\cdot}$) es la suma de todas las observaciones en el i-ésimo (j-ésimo) nivel de A (B) y $T_{\cdot\cdot k}$ es la suma de todas las observaciones en la k-ésima repetición. En forma similar, $T_{ij\cdot}$ es la suma de todas las observaciones en el tratamiento (i, j). Las definiciones correspondientes para las medias de la muestra deben ser aparentes.

Al reemplazar los parámetros en (12.12) con sus correspondientes estimadores, se tiene

$$(Y_{ijk} - \overline{Y}_{\cdot\cdot\cdot}) = (\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot\cdot\cdot}) + (\overline{Y}_{\cdot j\cdot} - \overline{Y}_{\cdot\cdot\cdot}) + (\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y}_{\cdot\cdot\cdot}) + (Y_{ijk} - \overline{Y}_{ij\cdot}).$$

Si se eleva al cuadrado la identidad con base en la muestra anterior y se suman sobre i, j y k, todos los términos que contienen productos cruzados se reducen a cero, y se tiene el siguiente resultado:

$$\sum_i \sum_j \sum_k (Y_{ijk} - \overline{Y}_{\cdot\cdot\cdot})^2 = nb \sum_i (\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot\cdot\cdot})^2 + na \sum_j (\overline{Y}_{\cdot j\cdot} - \overline{Y}_{\cdot\cdot\cdot})^2 + n \sum_i \sum_j (\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y}_{\cdot\cdot\cdot})^2 + \sum_i \sum_j \sum_k (Y_{ijk} - \overline{Y}_{ij\cdot})^2. \qquad (12.23)$$

En otras palabras, la suma total de cuadrados se separa en las sumas de cuadrados debidas: al factor A *(SCA)*, el factor B *(SCB)*, a la interacción entre A y B *(SCAB)* y a los errores *(SCE)*.

También puede escribirse el modelo (12.21) en términos de las desviaciones causadas por los tratamientos y el error aleatorio, es decir

$$(Y_{ijk} - \mu) = (\mu_{ij\cdot} - \mu) + (Y_{ijk} - \mu_{ij\cdot}). \qquad (12.24)$$

En esta forma, la desviación debida a los tratamientos abarca los efectos debidos a A, B y la interacción $A\ B$. Al sustituir en (12.24) las correspondientes estadísticas, se tiene

$$(Y_{ijk} - \overline{Y}_{\cdot\cdot\cdot}) = (\overline{Y}_{ij\cdot} - \overline{Y}_{\cdot\cdot\cdot}) + (Y_{ijk} - \overline{Y}_{ij\cdot}),$$

las que, al elevarse al cuadrado y sumar sobre i, j y k, dan como resultado

$$\sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{...})^2 = n \sum_i \sum_j (\bar{Y}_{ij\cdot} - \bar{Y}_{...})^2 + \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij\cdot})^2,$$

o

$$\text{STC} = \text{SCTR} + \text{SCE}. \tag{12.25}$$

De (12.23) se desprende que

$$\text{SCTR} = \text{SCA} + \text{SCB} + \text{SCAB}. \tag{12.26}$$

Puede demostrarse que, con base en (12.23), la descomposición del número de grados de libertad es la siguiente:

$$\text{gl(STC)} = \text{gl(SCA)} + \text{gl(SCB)} + \text{gl(SCAB)} + \text{gl(SCE)},$$

o

$$(nab - 1) = (a - 1) + (b - 1) + (a - 1)(b - 1) + ab(n - 1).$$

Para las suposiciones del modelo y la hipótesis de interés, SCA/σ^2, SCB/σ^2, SCAB/σ^2, y SCE/σ^2 son variables aleatorias independientes chi-cuadrada con $(a - 1)$, $(b - 1)$, $(a - 1)(b - 1)$ y $ab(n - 1)$ grados de libertad, para cada una. De acuerdo con lo anterior, la estadística de prueba para los efectos principales y de interacción son los cocientes entre los cuadrados medios, correspondientes y cuadrado medio del error y tienen una distribución F. Al igual que para los casos anteriores, una región crítica de tamaño α en el extremo superior de la región es la apropiada para cada caso. Puede observarse que el resultado anterior sigue siendo válido al examinar los valores esperados de los cuadrados medios. Para el caso de efectos fijos, estos valores son los siguientes:

$$E(\text{CME}) = \sigma^2,$$

$$E(\text{CMA}) = \sigma^2 + nb\frac{\sum \alpha_i^2}{a - 1},$$

$$E(\text{CMB}) = \sigma^2 + na\frac{\sum \beta_j^2}{b - 1},$$

$$E(\text{CMAB}) \quad {}^2 + n\frac{\sum\sum (\alpha\beta)_{ij}^2}{(a - 1)(b - 1)}.$$

Si no existe ninguna interacción entre A y B (es decir, si $(\alpha\beta)_{ij} = 0$ para toda i y j), entonces $CMAB$ y CME tienen el mismo valor esperado y los efectos son aditivos. Pero si el cociente $CMAB/CME$ tiene un valor suficientemente grande, esto sugeriría una interacción estadísticamente apreciable entre A y B y, por lo tanto, debe rechazarse la hipótesis nula. De manera similar si $\alpha_i = 0$ para toda i, CMA y CME tienen valores esperados iguales y no existe un efecto principal causado por A. Pero un cociente grande entre CMA y CME tiende a implicar que el efecto principal

atribuible a A es estadísticamente significativo. El mismo argumento es válido para el efecto principal de B.

En la tabla 12.14 se encuentra un resumen del análisis de varianza para un diseño factorial con dos factores. Aunque en la tabla se proporcionan fórmulas de cálculo para cada fuente de variación, la práctica usual para realizarlos a mano es calcular *SCT* mediante la fórmula que aparece en la tabla 12.14 y *SCTR* de la fórmula

$$\text{SCTR} = \frac{1}{n}\sum_i \sum_j T_{ij\cdot}^2 - \frac{T_{\cdots}^2}{nab}.$$

Entonces puede obtenerse *SCE* al emplear (12.25). A su vez, mediante el empleo de las fórmulas que aparecen en la tabla 12.14 se calculan *SCA* y *SCB*, y se obtiene *SCAB* con base en (12.26).

Ejemplo 12.5 Se llevó a cabo una investigación para determinar si pueden encontrarse diferencias apreciables en los salarios iniciales para contadores graduados con base en el sexo, localidad del lugar de trabajo o la interacción de los dos. El estudio se llevó a cabo en grandes ciudades del noroeste, el oeste medio y el oeste. Se piensa que será suficiente un arreglo factorial en un diseño completamente aleatorizado. Se decide emplear los salarios iniciales de cuatro personas para cada una de las seis combinaciones de tratamientos. Para asegurar que las unidades experimentales son homogéneas, se seleccionaron personas con antecedentes muy similares en la medida de lo posible. Tienen la misma edad y el mismo promedio de calificaciones durante sus estudios; ninguno tenía experiencia profesional y todos se graduaron en

TABLA 12.14 Tabla ANOVA para un experimento factorial con dos factores completamente aleatorizados

Fuente de variación	gl	SC	CM	*Estadística F*
Factor A	$a-1$	$\frac{1}{nb}\sum_i T_{i\cdot\cdot}^2 - \frac{T_{\cdots}^2}{nab}$	SCA $/(a-1)$	CMA/CME
Factor B	$b-1$	$\frac{1}{na}\sum_j T_{\cdot j\cdot}^2 - \frac{T_{\cdots}^2}{nab}$	SCB $/(b-1)$	CMB/CME
Interacción AB	$(a-1)(b-1)$	$\frac{1}{n}\sum_i\sum_j T_{ij\cdot}^2 - \frac{1}{nb}\sum_i T_{i\cdot\cdot}^2 - \frac{1}{na}\sum_j T_{\cdot j\cdot}^2 + \frac{T_{\cdots}^2}{nab}$	SCAB$/(a-1)(b-1)$	CMAB/CME
Error	$ab(n-1)$	$\sum_i\sum_j\sum_k Y_{ijk}^2 - \frac{1}{n}\sum_i\sum_j T_{ij\cdot}^2$	SCE $/ab(n-1)$	
Total	$nab-1$	$\sum_i\sum_j\sum_k Y_{ijk}^2 - \frac{T_{\cdots}^2}{nab}$		

TABLA 12.15 Salarios iniciales para contadores graduados (miles de dólares)

	Noroeste	*Oeste medio*	*Oeste*	*Totales*
Mujeres	15.2	14.9	16.2	
	16.8	16.2	15.9	
	15.5	15.6	16.8	
	14.9	15.3	15.8	
	$T_{11\cdot} = 62.4$	$T_{21\cdot} = 62.0$	$T_{31\cdot} = 64.7$	$T_{\cdot 1\cdot} = 189.1$
Hombres	18.1	17.8	18.4	
	16.3	18.2	16.8	
	17.2	18.1	17.5	
	17.9	17.6	18.7	
	$T_{12\cdot} = 69.5$	$T_{22\cdot} = 71.7$	$T_{32\cdot} = 71.4$	$T_{\cdot 2\cdot} = 212.6$
Totales	$T_{1\cdot\cdot} = 131.9$	$T_{2\cdot\cdot} = 133.7$	$T_{3\cdot\cdot} = 136.1$	$T_{\cdot\cdot\cdot} = 401.7$

universidades del mismo nivel académico. Con base en la información de la muestra proporcionada en la tabla 12.15, determínense cuáles efectos son estadísticamente apreciables.

Las sumas de interés aparecen en la tabla. Entonces

$$\text{STC} = 15.2^2 + 16.8^2 + \cdots + 18.7^2 - \frac{401.7^2}{24} = 32.8563,$$

$$\text{SCTR} = \frac{62.4^2 + 69.5^2 + \cdots + 71.4^2}{4} - \frac{401.7^2}{24} = 24.7838,$$

$$\text{SCE} = 32.8563 - 24.7838 = 8.0725.$$

De manera similar,

$$\text{SC(SEX)} = \frac{189.1^2 + 212.6^2}{12} - \frac{401.7^2}{24} = 23.0104,$$

$$\text{SC(LOC)} = \frac{131.9^2 + 133.7^2 + 136.1^2}{8} - \frac{401.7^2}{24} = 1.11.$$

De esta forma

$$\text{SC(LOC} \times \text{SEX)} = 24.7838 - 23.0104 - 1.11 = 0.6634.$$

La tabla del análisis de varianza se encuentra en la tabla 12.16. Con base en esta información, puede concluirse que el único efecto discernible estadísticamente en el salario inicial se debe al sexo del graduado.

Debe notarse que el método de Scheffé para comparar las medias del nivel del factor se extiende, en forma directa, a experimentos factoriales. También puede

TABLA 12.16 Tabla ANOVA para el ejemplo 12.5

Fuente de variación	gl	SC	CM	*Valor F*
Localidad	2	1.11	0.555	1.24
Sexo	1	23.0104	23.0104	51.31
Localidad × sexo	2	0.6634	0.3317	0.74
Error	18	8.0725	0.4485	
Total	23	32.8563	$f_{0.99,1,18} = 8.29$; $f_{0.99,2,18} = 6.01$	

efectuarse un análisis de residuos para los niveles de cada factor para verificar, entre otras cosas, la hipótesis de varianzas iguales. Los residuos se obtienen mediante el empleo de la relación

$$e_{ijk} = y_{ijk} - \bar{y}_{ij}.$$

En los casos que se han examinado hasta este momento, siempre se empleó el cuadrado medio del error como el denominador del cociente *F*. Sin embargo, para experimentos estadísticos que incluyen dos o más factores, lo anterior no siempre es válido. La estadística *F* apropiada para un análisis de varianza depende, en forma directa, de las esperanzas de los cuadrados medios de las fuentes de variación, las que a su vez dependen de si se consideran a los efectos correspondientes como fijos o aleatorios.

Para experimentos factoriales con dos factores surgen tres situaciones distintas: *a*) los niveles de ambos factores son de efectos fijos; *b*) los niveles de ambos factores son de efectos aleatorios, o *c*) los niveles de un factor son de efectos fijos mientras que los del otro son de efectos aleatorios. Ya se ha analizado la primera posiblidad. Para las otras dos, los valores esperados de los cuadrados medios tanto para el modelo de efectos aleatorios como para el modelo de efectos mixtos se proporcionan en la tabla 12.17.

TABLA 12.17 Esperanzas de cuadrados medios para un factorial con dos factores: modelos de efectos aleatorios o de efectos mixtos

	Efectos aleatorios (A y B aleatorios)		*Efectos mixtos (A fijo, B aleatorio)*	
Fuente	ECM	*Estadística F*	ECM	*Estadística F*
A	$\sigma^2 + n\sigma^2_{\alpha\beta} + nb\sigma^2_\alpha$	CMA/CMAB	$\sigma^2 + n\sigma^2_{\alpha\beta} + nb\dfrac{\sum \alpha_i^2}{(a-1)}$	CMA/CMAB
B	$\sigma^2 + n\sigma^2_{\alpha\beta} + na\sigma^2_\beta$	CMB/CMAB	$\sigma^2 + na\sigma^2_\beta$	CMB/CME
AB	$\sigma^2 + n\sigma^2_{\alpha\beta}$	CMAB/CME	$\sigma^2 + n\sigma^2_{\alpha\beta}$	CMAB/CME
Error	σ^2		σ^2	

Con base en el material de este capítulo, el procedimiento que se ha empleado para construir la estadística de prueba es comparar dos cuadrados medios que, bajo la hipótesis nula, tengan el mismo valor esperado, y bajo la hipótesis alternativa, el cuadrado medio del numerador tenga un valor esperado mucho más grande que el del denominador correspondiente. Si la hipótesis nula es cierta, la estadística tiene una distribución F con un número apropiado de grados de libertad. Con esto en mente, los cocientes de cuadrados medios indicados en la tabla 12.17 deben ser ya evidentes. Por ejemplo, considérese el caso de efectos aleatorios y, en particular, la hipótesis nula de que no existe variación alguna entre todos los posibles niveles de A; esto es, H_0: $\sigma_\alpha^2 = 0$. Si H_0 es cierta, entonces $E(CMA) = \sigma^2 + n\sigma_{\alpha\beta}^2$, donde $\sigma_{\alpha\beta}^2$ denota la varianza causada por la interacción entre A y B. Pero este valor esperado es el mismo sólo para $E\ (CMAB)$ y no para $E\ (CME)$ bajo H_0. Por otro lado, si H_0 es falsa, $E\ (CMA)$ es mayor que $E\ (CMAB)$. De acuerdo con lo anterior, la estadística de prueba apropiada para H_0 es $CMA/CMAB$.

Debe recordarse que en experimentos factoriales, el cuadrado medio del error será el denominador en el cociente de cuadrados medios para todos los efectos principales y de interacción, sólo si los niveles de todos los factores son de efectos fijos. De esta forma, en la fase de diseño de un experimento estadístico es muy importante la selección de los niveles del factor, ya que tienen una influencia directa en el análisis.

Referencias

1. W. G. Cochran and G. M. Cox, *Experimental designs*, 2nd ed., Wiley, New York, 1957.
2. R. C. Hicks, *Fundamental concepts in the design of experiments*, 2nd ed., Holt, Rinehart and Winston, New York, 1973.
3. R. L. Horton, *The general linear model*, McGraw-Hill, New York, 1978.
4. R. E. Kirk, *Experimental design: Procedures for the behavioral sciences*, Brooks/Cole, Belmont, Calif., 1968.
5. J. Neter and W. Wasserman, *Applied linear statistical models*, Richard D. Irwin, Homewood, Ill., 1974.
6. H. Scheffé, *Analysis of variance*, Wiley, New York, 1953.
7. H. Scheffé, *A method for judging all contrasts in the analysis of variance*, Biometrika **40** (1953), 87–104.

Ejercicios

12.1. Suponga que se asigna al lector la responsabilidad de investigar en una fábrica el efecto que pueden tener diferentes cambios en la semana de 40 horas de trabajo, sobre la productividad promedio en una gran fábrica. En forma específica, se desean comparar cinco días a la semana, 4 días a la semana y 3⅓-días a la semana. Describa con gran detalle su propuesta de diseño estadístico. Asegúrese de identificar los tratamientos, las unidades experimentales y otros factores importantes para llevar a cabo la investigación.

12.2. Las estadísticas para accidentes indican que alrededor de dos terceras partes de los accidentes automovilísticos de consecuencias fatales en Estados Unidos son causados por conductores en estado de ebriedad. Supoga que usted es comisionado para investigar el

grado en el que el alcohol afecta la habilidad de las personas para desempeñar funciones de rutina al conducir un automóvil. Descríbase con gran detalle un diseño estadístico para lograr esta tarea e indíquese cómo debe llevarse a cabo este experimento.

12.3. Una compañía de seguros desea determinar si existen diferencias discernibles en el número de días promedio que los pacientes que padecen una misma enfermedad permanecen en cuatro grandes hospitales de cierta área metropolitana. La compañía también está interesada en detectar cualquier efecto debido al sexo de los pacientes. Descríbase con detalle un diseño estadístico para lograr este objetivo. Asegúrese de identificar la naturaleza de cada factor, ya sea como de efecto fijo o aleatorio; escríbase el modelo y establézcase la hipótesis por probar.

12.4. Una operación de llenado tiene tres máquinas idénticas que se ajustan para vaciar una cantidad específica de un producto en recipientes de igual tamaño. Con el propósito de verificar la igualdad de las cantidades promedio vaciadas por cada máquina, se toman muestras aleatorias, en forma periódica, de cada una. Para un periodo particular, se observaron los datos que aparecen en la tabla 12.18.

TABLA 12.18 Datos de la muestra para el ejercicio 12.4

	Máquina	
A	B	C
16	18	19
15	19	20
15	19	18
14	20	20
	19	19
	19	

a) Calcúlese $y_{ij} - \bar{y}_{..}$ y verifíquese que la suma de estas desviaciones para toda i y j es cero.

b) Estímese τ_j para toda j, y verifíquese que la suma de $n_j(\bar{y}_{.j} - \bar{y}_{..})$ sobre todas las j es cero.

c) Calcúlese, en forma directa, cada una de las tres sumas de cuadrados dadas en la expresión 12.8 para verificar que $STC = SCTR + SCE$.

d) ¿Existen algunas diferencias estadísticamente significativas en las cantidades promedio vaciadas por las tres máquinas? Empleése $\alpha = 0.05$.

12.5. En el ejercicio 12.4, supóngase que se divide cada observación entre 10. Demuéstrese si esta operación tiene algún efecto con las respuestas a las partes *c* y *d*.

12.6. Para el ejercicio 12.4, constrúyanse constrastes a su elección y empléese el método de Scheffé para determinar si éstos son estadísticamente significativos.

12.7. Se pide a un laboratorio de prueba independiente que compare la durabilidad de cuatro diferentes marcas de pelotas de golf. El laboratorio propone un experimento en el que se seleccionan, en forma aleatoria ocho pelotas por cada fabricante y se ponen en una máquina que golpea cada pelota con una fuerza constante. La medición de interés es el número de veces que la máquina golpea la pelota antes de que su recubrimiento externo se rompa. En la tabla 12.19 se encuentra la información que se obtuvo al llevar a cabo el experimento.

TABLA 12.19 Datos de la muestra para el ejercicio 12.7

	Marca		
A	B	C	D
205	242	237	212
229	253	259	244
238	226	265	229
214	219	229	272
242	251	218	255
225	212	262	233
209	224	242	224
204	247	234	245

a) ¿Existe alguna razón para creer que la durabilidad promedio es diferente para cada una de las cuatro marcas? Úsese $\alpha = 0.05$.
b) ¿Existe alguna razón para dudar de la suposición de que las varianzas de los errores son iguales?

12.8. Para determinar si existen diferencias en la cosecha promedio de tres variedades de maíz, se dividió en tres partes iguales un área para siembra. A su vez, cada una de estas partes se subdivide en otras cinco iguales entre sí, y se siembra cada una con una variedad de maíz. En el momento de la cosecha, la medición de interés es el número de toneladas por acre. La tabla 12.20 es una tabla de análisis de varianza incompleta para este problema.

TABLA 12.20. Tabla parcial ANOVA para el ejercicio 12.8

Fuente	gl	SC	CM	*Valor F*
Tratamientos		64		
Error				
Total		100		

a) Escríbase el modelo para este problema.
b) ¿Se está satisfecho con las suposiciones? Hágase un comentario.
c) Establézcase la hipótesis nula por probar.
d) Complétese la tabla ANOVA y determínese si puede rechazarse la hipótesis nula para un nivel $\alpha = 0.01$.

12.9. Se desea determinar si la cantidad de carbón empleado en la fabricación de acero tiene algún efecto en la resistencia a la tensión de éste. Se investigaron cinco diferentes porcentajes de carbón: 0.2, 0.3, 0.4, 0.5 y 0.6%. Para cada porcentaje de carbón se seleccionaron, en forma aleatoria del mismo lote, cinco muestras de acero y se midieron las resistencias a la tensión. Se obtuvo la información que se muestra en la tabla 12.21, donde la tensión se encuentra en kilogramos por centímetro cuadrado.

a) Con base en esta información, determínese si el porcentaje de carbón tiene un efecto estadísticamente significativo sobre la resistencia a la tensión del acero. Úsese $\alpha = 0.01$.
b) Si la respuesta a la parte *a* es afirmativa, propónganse los contrastes relevantes y pruébese su significancia estadística.

TABLA 12.21 Datos de la muestra para el ejercicio 12.9

		Contenido de carbón		
0.2%	0.3%	0.4%	0.5%	0.6%
1240	1420	1480	1610	1700
1350	1510	1470	1590	1790
1390	1410	1520	1580	1740
1280	1530	1540	1630	1810
1320	1470	1510	1560	1730

12.10. En el ejercicio 12.9, ¿existe alguna razón para dudar de la suposición de varianzas iguales?

12.11. Se seleccionó una muestra al azar de un número de presidentes de compañías, en cuatro diferentes áreas geográficas de Estados Unidos, con el propósito de determinar si el área tiene algún efecto sobre los ingresos anuales de estos altos ejecutivos. Se observaron los salarios anuales que se muestran en la tabla 12.22. Con la información dada, proporciónese un argumento, ya sea en contra o a favor, de si debe utilizarse la técnica del análisis de varianza para determinar si el área tiene algún efecto sobre el ingreso anual. Trátese de dar un apoyo sustancial en cualquiera de los dos casos.

TABLA 12.22 Datos de la muestra para el ejercicio 12.11 (miles de dólares)

	Área		
Noreste	*Oeste medio*	*Sureste*	*Oeste*
140	93	78	85
125	135	112	72
95	68	57	97
110	53	97	105
59	115	52	62

12.12. En una planta industrial se desea determinar si diferentes trabajadores con el mismo nivel de habilidad tienen algún efecto sobre el número de unidades que se espera que produzcan durante un periodo fijo. Se lleva a cabo un experimento en el que se seleccionan al azar cinco trabajadores y se observa el número de unidades que cada uno produce en seis periodos con la misma duración, produciéndose los resultados que se encuentran en la tabla 12.23.

TABLA 12.23 Datos de la muestra para el ejercicio 12.12

		Trabajador		
1	2	3	4	5
45	52	39	57	48
47	55	37	49	44
43	58	46	52	55
48	49	45	50	53
50	47	42	48	49
44	57	41	55	52

a) Escríbase el modelo para este problema y explíquese cada término.
b) Establézcase la hipótesis nula por probar.
c) Determínese si puede rechazarse la hipótesis nula para un nivel $\alpha = 0.05$.
d) ¿Qué fracción de la varianza en el número de unidades producidas es atribuible a diferencias entre los trabajadores?

12.13. Desde el incremento en los precios de la gasolina se han desarrollado varios dispositivos, los cuales se colocan en los carburadores de los automóviles, con el propósito de aumentar el rendimiento de éstos. Una empresa selecciona tres de los dispositivos más populares para someterlos a prueba. La empresa desea compararlos con los carburadores estándar, con el propósito de determinar si existe un incremento apreciable de millas por galón de gasolina con el uso de estos dispositivos. La compañía selecciona cinco tipos de automóviles para el experimento. Para controlar la variación, se planea utilizar el mismo conductor para todo el experimento.

TABLA 12.24 Datos de la muestra para el ejercicio 12.13 (millas por galón)

Automóvil	*Carburador estándar*	*Dispositivo* A	*Dispositivo* B	*Dispositivo* C
1	18.2	18.9	19.1	20.4
2	27.4	27.9	28.1	29.9
3	35.2	34.9	35.8	38.2
4	14.8	15.2	14.9	17.3
5	25.4	24.8	25.6	26.9

a) Hágase un bosquejo del plan específico para realizar este experimento.
b) Supóngase que se observan los datos que se encuentran en la tabla 12.24. Escríbase el modelo y establézcase la hipótesis nula por probar. ¿Puede rechazarse la hipótesis nula para un nivel $\alpha = 0.05$.
c) Si se rechaza la hipótesis nula de la parte *b*, constrúyanse por lo menos dos contrastes relevantes y pruébese su significancia estadística.

12.14. En el ejercicio 12.13, supóngase que no se ha considerado el automóvil como una fuente viable de variación en el rendimiento observado y muéstrese si esta omisión tiene algún efecto con la respuesta a la parte *b*.

12.15. Los cigarrillos producen cantidades apreciables de monóxido de carbono. Cuando se inhala el humo del cigarrillo, el monóxido de carbono se combina con la hemoglobina para formar carboxihemoglobina. En un estudio reciente,* los investigadores deseaban determinar si una concentración apreciable de carboxihemoglobina reduce la tolerancia al ejercicio en aquellos pacientes que sufren de bronquitis crónica y enfisema. Se seleccionaron siete** de estos pacientes y, en un ambiente controlado, se les pidió que caminaran durante 12 minutos respirando una de las siguientes cuatro mezclas gaseosas: aire, oxígeno, aire más monóxido de carbono (CO) u oxígeno más monóxido de carbono. La cantidad de monóxido de carbono respirado fue suficiente para elevar la concentración de carboxihemoglobina de cada sujeto en 9%. Para controlar el consumo de monóxido de carbono, se pidió a los siete fumadores que dejaran de fumar 12

*P. M. A. Calverly, R. J. E. Leggett, and D. C. Flenley, *Carbon monoxide and exercise tolerance in chronic bronchitis and emphysema*, Brit. Med. J. **283** (1981), 877–880.

** El estudio completo se llevó a cabo con 15 sujetos.

horas antes del experimento. Los datos que figuran el la tabla 12.25 representan las distancias caminadas por los sujetos en 12 minutos para cada condición.

TABLA 12.25 Datos de la muestra para el ejercicio 12.15 (en litros)

	Mezcla gaseosa			
Sujeto	*Aire*	*Oxígeno*	*Aire* + CO	*Oxígeno* + CO
1	835	874	750	854
2	787	827	755	829
3	724	738	698	726
4	336	378	210	279
5	252	315	168	336
6	560	672	558	642
7	336	341	260	336

a) Escríbase el modelo para este problema.

b) ¿Puede rechazarse la hipótesis nula de que no existe algún efecto, debido a la mezcla de gas, en la distancia caminada durante el lapso de 12 minutos para un nivel de $\alpha = 0.05$?

c) Llévese a cabo una prueba *F* conservadora para la hipótesis nula. ¿Es la conclusión diferente a la de la parte *b*?

d) Si la respuesta a la parte *b* es sí, constrúyanse los contrastes pertinentes y empléese el método de Scheffé para determinar si éstos son estadísticamente significativos.

12.16. Se desea determinar si existen diferencias apreciables en los precios promedio entre cuatro grandes supermercados en una ciudad dada. De los artículos de la misma marca que se venden con regularidad, se seleccionan al azar 10 y se observan sus precios unitarios en cada supermercado. Se obtiene la información que figura en la tabla 12.26.

a) Escríbase el modelo para este problema.

b) Establézcase una hipótesis nula apropiada y determínese si ésta puede rechazarse para un nivel de $\alpha = 0.01$.

c) Determínense todos los residuos y hágase la gráfica de éstos para cada tratamiento y para cada bloque. Hágase un comentario sobre sus resultados.

TABLA 12.26 Datos de la muestra para el ejercicio 12.16 (en dólares)

	Supermercado			
Artículo	A	B	C	D
1	3.29	3.42	3.27	3.35
2	0.59	0.65	0.59	0.60
3	1.25	1.29	1.25	1.27
4	4.35	4.59	4.29	4.49
5	0.89	0.95	0.89	0.89
6	1.85	1.79	1.89	1.89
7	0.95	0.89	0.89	0.90
8	0.75	0.79	0.69	0.79
9	2.35	2.35	2.39	2.39
10	1.49	1.55	1.55	1.49

12.17. En el ejemplo que sirvió como introducción en la sección 12.6, supóngase que se seleccionan en forma aleatoria 12 componentes del mismo lote y en grupos de tres se asig-

nan a las cuatro combinaciones de hornos y temperaturas. Los tiempos de duración de los componentes se encuentran en la tabla 12.27.

TABLA 12.27 Datos de la muestra para el ejercicio 12.17 (en horas)

	O_1	O_2
T_1	6.29	5.95
	6.38	6.05
	6.25	5.89
T_2	5.80	6.32
	5.92	6.44
	5.78	6.29

a) Escríbase el modelo apropiado para este problema.
b) Establézcase la hipótesis por probar.
c) Determínese la tabla del análisis de varianza y obténganse conclusiones apropiadas. Empléese $\alpha = 0.05$.

12.18. En el ejercicio 12.3, supóngase que se obtuvo la información proporcionada en la tabla 12.28 para pacientes seleccionados al azar, que padecen la misma enfermedad.

TABLA 12.28 Datos de la muestra para el ejercicio 12.18. Duración de la hospitalización en días en cuatro hospitales.

	Hospital A	*Hospital* B	*Hospital* C	*Hospital* D
Hombres	7	9	10	6
	10	9	8	7
	8	12	12	6
	11	14	13	9
Mujeres	9	11	13	8
	12	12	11	9
	12	14	14	8
	11	13	14	10

a) Determínese qué efectos son estadísticamente discernibles a un nivel de $\alpha = 0.01$.
b) Determínense todos los residuos y hágase la gráfica de éstos para cada hospital. ¿Qué conclusión puede dar?

12.19. El objetivo de un experimento de agricultura fue determinar si existían diferencias apreciables en la cantidad de trigo cosechado, de entre cuatro variedades y tres tipos de fertilizantes. Para el experimento se encontró una área muy grande de siembra en la que las condiciones del suelo eran, prácticamente, homogéneas. El área fue dividida en 12 zonas de igual tamaño para las 12 combinaciones de variedad de trigo y tipo de fertilizante. Para medir el error experimental, cada zona se dividió a su vez en cuatro y cada una de éstas recibió el mismo tratamiento. Las tres clases de fertilizante se seleccionaron, en forma aleatoria, de entre un número relativamente grande de fertilizantes, pero el interés no se extendió más allá de las cuatro variedades de trigo seleccionadas para el experimento. En el momento de la cosecha se observaron los datos que aparecen en la tabla 12.29.

TABLA 12.29 Datos de la muestra para el ejercicio 12.19 (toneladas por acre)

	Variedad de trigo			
Fertilizante	A	B	C	D
1	35	45	24	55
	26	39	23	48
	38	39	36	39
	20	43	29	49
2	55	64	58	68
	44	57	74	61
	68	62	49	60
	64	61	69	75
3	97	93	89	82
	89	91	98	78
	92	82	85	89
	99	98	87	92

a) Escríbase el modelo apropiado para este problema.
b) Establézcase la hipótesis nula por probar.
c) Determínese la tabla de análisis de varianza y obténganse las conclusiones apropiadas. Úsese $\alpha = 0.05$.

12.20. En el ejercicio 12.19, ¿Cómo puede cambiar la respuesta a la parte *c*, si

a) ¿Se supone que las variedades son de efectos aleatorios, y los tipos de fertilizante son de efectos fijos?
b) ¿Se supone que ambos son de efectos fijos?
c) ¿Se supone que ambos son de efectos aleatorios?

CAPÍTULO TRECE

Análisis de regresión: el modelo lineal simple

13.1 Introducción

En el capítulo anterior se desarrollaron los criterios básicos para el diseño estadístico de experimentos. En este capítulo se examinarán las asociaciones cuantitativas entre un número de variables, lo que en la terminología estadística se conoce como ***análisis de regresión***.

Aunque en muchas disciplinas se están realizando experimentos diseñados en forma estadística, la precisión en la comparación que en forma general se requiere, evita el empleo de estos diseños en muchas situaciones. Investigar el efecto simultáneo de varios factores con base en las técnicas del análisis de varianza requiere de la suposición de que los datos se han colectado en arreglos balanceados y que se llevaron a efecto los procedimientos de aleatorización adecuados. En forma obvia, lo anterior es deseable si puede cumplirse, pero muchas veces es impráctico. En realidad, a lo que en general se enfrenta el experimento es a un conjunto de datos que de manera común, no espera que hayan sido observados bajo condiciones estrictamente controladas y los que, salvo en ciertas ocasiones, no contienen ninguna réplica real que permita una estimación apropiada del error experimental. Bajo estas condiciones, los métodos más apropiados son el de mínimos cuadrados y el análisis de regresión, y no los del análisis de varianza.

El propósito de este capítulo radica en proporcionar los conceptos y metodología básicos para extraer de grandes cantidades de datos las características principales de una relación que no es evidente. De manera específica, se examinarán técnicas que permitan ajustar una ecuación de algún tipo al conjunto de datos dado, con el propósito de obtener una ecuación empírica de predicción razonablemente precisa y que proporcione un modelo teórico que no está disponible. Se supondrá la existencia de un conjunto de n mediciones $y_1, y_2, \ldots, y_n$ de una variable respuesta Y, las cuales se han observado bajo un conjunto de condiciones experimentales $(x_1, x_2, \ldots, x_k)$ que representan los valores de k variables de predicción. El interés recae en determinar una función matemática sencilla, por ejemplo un polinomio que describa, en

forma razonable, el comportamiento de la variable respuesta, dados los valores de las variables de predicción. Nótese que la ecuación que se obtiene por esta forma puede tener algunas limitaciones con respecto a su interpretación física; sin embargo, en un medio empírico, será muy útil si puede proporcionar una adecuada capacidad de predicción para la respuesta en el interior de una región especificada de las variables de predicción.

A pesar de que no se encuentra problema alguno con las designaciones comunes de variable dependiente e independiente para Y y x, respectivamente, se preferirá denominarlas como variable de *respuesta* y de *predicción*, ya que en la regresión sólo puede *asociarse* un valor de Y con uno de predicción x; no es posible establecer una relación causa-efecto entre la Y y las x. Algunos ejemplos proporcionarán una idea del por qué obtener una relación causa-efecto se encuentra más allá del alcance del análisis de regresión. De manera obvia, existe una relación entre la altura y el peso de los seres humanos, pero ¿implica esta relación, por ejemplo, que pueda cambiar la altura de una persona si se modifica su peso? También se tiene una relación entre la cantidad de gas bruto que se consume en cierta área de alguna ciudad y la temperatura atmosférica promedio, pero ¿significa esto que es posible aumentar la temperatura mediante la reducción del consumo de gas? También puede existir alguna relación entre un factor económico en particular y un ciclo financiero, pero ¿implica lo anterior que el factor económico "causa" el ciclo financiero?

La esencia de los ejemplos anteriores está en el hecho de que el análisis de regresión sólo descubre una asociación entre la variable de respuesta y las variables de predicción, en lugar de detectar una relación causa-efecto. La causalidad implica que un cambio en las x causará uno correspondiente en la variable respuesta. Por ejemplo, cuando se calienta un metal éste se expande; en este caso no existe ninguna duda de que establecer una relación causa-efecto es muy importante. Pero en forma desafortunada, lo anterior generalmente no puede llevarse a cabo con base en un análisis estadístico, a menos que se efectúe un experimento rigurosamente controlado. Un ejemplo de lo anterior, es la relación que existe entre fumar y el cáncer pulmonar. La evidencia que se tiene resulta abrumadora con respecto a que el fumador crónico (predicción) está estadísticamente asociado con una alta incidencia de cáncer pulmonar (respuesta). La industria cigarrera argumenta, en contra de estos hallazgos, que todavía no existe una relación de tipo causal entre fumar mucho y la incidencia de cáncer pulmonar.

El enfoque que se utilizará en este capítulo, así como en el siguiente, se limitará a establecer el grado de asociación que existe entre variables, sin tomar en cuenta la noción de causalidad. En este capítulo se examinarán los fundamentos del análisis de regresión para el modelo con una sola variable de predicción. En el capítulo 14 se estudiará lo que se conoce como el *modelo lineal general* en el que se supone que una respuesta dada es una función de varias variables de predicción.

13.2 El significado de la regresión y suposiciones básicas

Si los métodos de regresión son tan útiles en la práctica, debe comprenderse su significado y las suposiciones bajo las cuales se han desarrollado. Las técnicas de regre-

sión proporcionan medios legítimos a través de los cuales pueden establecerse asociaciones entre las variables de interés en las cuales la relación usual no es causal. La palabra "regresión" se usó por primera vez en este contexto por Francis Galton (1822-1911) en sus estudios biológicos sobre la herencia. En ellos se notó que las características promedio de la siguiente generación de un grupo en particular tendían a moverse en la dirección de las características promedio de la población general, más que hacia las de la generación previa de ese grupo. Esta tendencia fue referida como una regresión hacia la media de la población.

De manera básica, la regresión tiene dos significados: uno surge de la distribución conjunta de probabilidad de dos variables aleatorias; el otro es empírico y nace de la necesidad de ajustar alguna función a un conjunto de datos. Para ilustrar el primer significado se tratará de predecir el salario anual de un profesionista dado el número de años que han transcurrido desde su graduación. Sea X el número de años y Y el salario anual. Debe ser obvio que para un valor dado de x es imposible predecir, de manera exacta, el salario anual de una persona en particular. Sin embargo, es posible predecir el salario promedio de todos aquellos individuos para los que el número de años x que han transcurrido desde su graduación es el mismo. En otras palabras, para cada valor de x existe una distribución de ingresos anuales y lo que se busca es la media de esa distribución, dado x. La gráfica de la media condicional $E(Y|x)$ como una función de x recibe el nombre de *curva de regresión* de Y sobre X. De esta forma, si $f(x, y)$ es la función de densidad conjunta de probabilidad de X y Y, y si $f(y|x)$ es la función de densidad condicional de Y dado x, se define la curva de regresión como

$$E(Y|x) = \int_{-\infty}^{\infty} yf(y|x)\,dy.$$

Ejemplo 13.1 Considérese la función de densidad conjunta de probabilidad dada por

$$f(x, y) = \begin{cases} 2x & 0 < x < y < 1, \\ 0 & \text{para cualquier otro valor} \end{cases}$$

Obténgase la curva de regresión de Y sobre X.

Dado que

$$f(y|x) = f(x, y)/f_X(x),$$

entonces

$$f_X(x) = \int_Y f(x, y)\,dy = \int_x^1 2x\,dy = 2x(1 - x),$$

y

$$f(y|x) = \frac{2x}{2x(1 - x)} = \frac{1}{1 - x}.$$

Por lo tanto, la curva de regresión es

$$E(Y|x) = \int_x^1 (1 - x)^{-1} y\, dy = (1 + x)/2,$$

la cual es una línea recta con pendiente e intersección igual a 1/2.

El segundo significado de la regresión es mucho más práctico que el primero. En él no se tienen los elementos necesarios para determinar la curva de regresión tal como se hizo en el ejemplo 13.1. No obstante, dado un conjunto de datos, puede asumirse una forma funcional para la curva de regresión y entonces tratar de ajustar ésta a los datos. En estas situaciones, la variable respuesta es una variable aleatoria cuyos valores se observan mediante la selección de los valores de las variables de predicción en un intervalo de interés. Por lo tanto, las variables de predicción no se consideran como variables aleatorias, sino que éstas son un conjunto de valores fijos que representan los puntos de observación para la variable respuesta. El modelo de regresión propuesto debe ser relativamente sencillo y deberá contener pocos parámetros. Un procedimiento muy útil para la selección inicial cuando se tiene sólo una variable de predicción es graficar la variable respuesta contra la variable de predicción. Si esta gráfica revela una tendencia lineal, deberá suponerse un modelo de regresión lineal. Si es evidente alguna curvatura, deberá suponerse un modelo cuadrático o de mayor grado para ajustarse a los datos.

Una vez que se ha seleccionado el modelo, la siguiente tarea es la de obtener estimaciones para los parámetros que intervienen en el mismo. Una técnica muy aceptada para este propósito es el *método de mínimos cuadrados (MC).* Este método encuentra las estimaciones para los parámetros en la ecuación seleccionada mediante la minimización de la suma de los cuadrados de las diferencias entre los valores observados de la variable respuesta y de aquéllos proporcionados por la ecuación de predicción. Estos valores se conocen como los estimadores por mínimos cuadrados (*EMC*) de los parámetros. Los estimadores por mínimos cuadrados poseen ciertas propiedades deseables, pero para determinarlas es necesario formular las siguientes suposiciones:

1. Se ha seleccionado la forma correcta de la ecuación de regresión. Esto implica que cualquier variabilidad en la variable respuesta que no pueda explicarse mediante el empleo de la ecuación de regresión, se debe a un error aleatorio. Por ejemplo, se sabe que la distancia d que recorre un objeto en un tiempo t, está dada por la siguiente relación

$$d = \beta_0 + \beta_1 t,$$

donde β_1 es la velocidad promedio y β_0 es la posición del objeto para $t = 0$. Si no fuese posible medir d en forma precisa para un valor dado de t, pero se observó un valor

$$y = d + \varepsilon,$$

donde ε es el error aleatorio, se ha seleccionado la forma correcta de la ecuación de regresión y el problema se reduce a estimar los valores de β_0 y β_1. Sin embargo, rara es la vez que el problema resulta ser tan sencillo.

Por ejemplo, si se tiene interés en predecir la cantidad de ozono que se encuentra en la estratósfera, como una función de los niveles de concentración de los constituyentes químicos de ésta en cierto momento del día, la ecuación por seleccionar será, en primera instancia, una conjetura. El error no puede considerarse como puramente aleatorio ya que pueden existir variaciones sistemáticas por causa de errores en el modelo. Algunos de los valores de la variable de respuesta proporcionados por la ecuación de predicción estarán sesgados ya que las estimaciones de los parámetros también se encuentran sesgadas.

2. Los datos que se observan son comunes en el sentido en que constituyen una muestra representativa de un medio acerca del cual el investigador desea generalizar. Si el investigador sabe que los datos no son representativos, el comportamiento general del mecanismo puede encontrarse más allá del alcance de los datos.

3. Los valores observados de la variable respuesta no se encuentran estadísticamente correlacionados. Se supone que cada valor observado está constituido por un valor real y una componente aleatoria. La componente aleatoria consiste en una variable aleatoria no observable; entonces la covarianza entre cualesquiera dos observaciones Y_i y Y_j, o entre los correspondientes errores aleatorios ε_i y ε_j, es cero para toda $\neq j$.

4. Para toda $i = 1, 2 \ldots n$, la media de ε_i es cero y la varianza de ε_i es σ^2. Esta última recibe el nombre de *varianza del error* y, generalmente, no es conocida. Dado que las variables de predicción no son variables aleatorias, la varianza de Y_i también es σ^2 para toda i y de esta forma es independiente del punto de observación. Si no es posible formular la suposición de que la varianza es constante para las observaciones de la variable respuesta, generalmente se emplea el método de mínimos cuadrados con factores de peso. Este tema se estudiará con cierto detalle en el capítulo 14.

5. Los puntos de observación o los valores de las variables de predicción son fijos o se seleccionan con anticipación y se miden sin error. Para muchas situaciones prácticas, ambas condiciones no se cumplen. Afortunadamente, el método de mínimos cuadrados sigue siendo válido siempre y cuando los errores en los valores de las x sean pequeños al compararse con los errores aleatorios y dado que éstos no dependan de los parámetros del modelo.

A manera de comentario final sobre las suposiciones del procedimiento *MC*, se considerarán sólo mínimos cuadrados lineales, donde la palabra "lineal" significa que el modelo seleccionado es lineal en los parámetros. La frase "lineal en los parámetros" significa que ningún parámetro en el modelo aparece como un exponente o es multiplicado por o dividido entre cualquier otro parámetro. Por ejemplo, los modelos

$$Y = \beta_0 + \beta_1 x + \varepsilon,$$

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon,$$

$$Y = \beta_0 + \beta_1 \ln(x) + \varepsilon,$$

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

son lineales en los parámetros β_0, β_1, β_2, y β_3, pero el modelo

$$Y = \beta_0 \exp(\beta_1 x) + \varepsilon$$

no lo es debido a que el parámetro β_1 aparece como un exponente.

13.3 Estimación por mínimos cuadrados para el modelo lineal simple

En esta sección se estudiará la estimación por mínimos cuadrados para el modelo lineal simple en el que sólo se tiene una variable de predicción, y se supone una ecuación de regresión lineal. Por ejemplo, los estudiantes universitarios que aprenden más rápido tienen mejores calificaciones promedio *(CP)* y por lo tanto, mejores oportunidades de obtener buenos empleos después de graduarse. Supóngase que los datos que se encuentran en la tabla 13.1 representan las calificaciones promedio de 15 recién graduados y sus correspondientes salarios iniciales.

Para este ejemplo, la variable respuesta es el salario inicial y la variable de predicción potencial es la calificación promedio. Estas últimas se seleccionaron de tal manera que reflejen un amplio intervalo. Se desea determinar una ecuación de regresión para el salario inicial promedio como una función de la calificación promedio. Dado que se ha propuesto sólo una variable de predicción, graficar los datos puede ser útil en la selección inicial de un modelo de regresión. La gráfica de los salarios iniciales contra las calificaciones promedio se muestra en la figura 13.1. Debe notarse que esta gráfica fue realizada por un paquete estadístico para computadora conocido como "minitab". Aunque no es tan sofisticado como SAS, Minitab es muy fácil de usar y se recomienda para llevar a cabo análisis preliminares de regresión, entre otras aplicaciones.

TABLA 13.1 Datos de la muestra para un modelo lineal simple (miles de dólares)

CP	*Salario inicial*
2.95	18.5
3.20	20.0
3.40	21.1
3.60	22.4
3.20	21.2
2.85	15.0
3.10	18.0
2.85	18.8
3.05	15.7
2.70	14.4
2.75	15.5
3.10	17.2
3.15	19.0
2.95	17.2
2.75	16.8

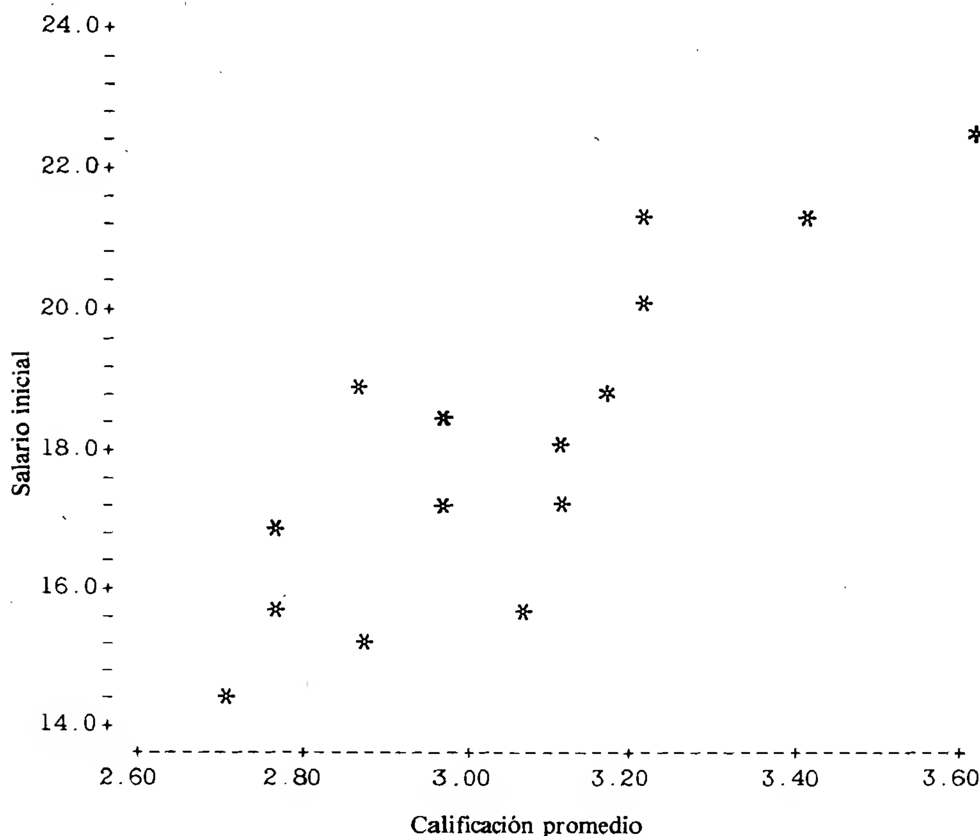


TABLA 13.1 Salario inicial contra calificación promedio

A pesar de que esta gráfica muestra una gran dispersión,* se observa una tendencia lineal. De acuerdo con lo anterior se supondrá un modelo de la forma

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad i = 1, 2, \ldots, n, \tag{13.1}$$

donde Y_i es la i-ésima observación de la variable respuesta, la cual corresponde al i-ésimo valor x_i de la variable de predicción, ε_i es el error aleatorio no observable asociado con Y_i; y β_0 y β_1 son los parámetros desconocidos que representan la intersección y la pendiente, respectivamente. La expresión (13.1) se conoce como *modelo lineal simple,* debido a que es lineal en los parámetros y se tiene sólo una variable de predicción.

Cada observación Y_i es una variable aleatoria que es la suma de dos componentes; el término no aleatorio $\beta_0 + \beta_1 x_i$, y la componente aleatoria ε_i. Si ε_i fuera un

* Por esta razón, este tipo de gráfica se conoce como gráfica de dispersión.

valor igual a cero, la observación Y_i se encontraría precisamente sobre la línea de regresión $\beta_0 + \beta_1 x_i$. Por lo tanto, ε_i es la distancia vertical de la observación a la línea de regresión. Dado que se supone

$$E(\varepsilon_i) = 0, \quad Var(\varepsilon_i) = \sigma^2 \qquad i = 1, 2, \ldots, n,$$

y

$$Cov(\varepsilon_i, \varepsilon_j) = 0 \qquad i \neq j;$$

entonces

$$E(Y_i) = E(\beta_0 + \beta_1 x_i + \varepsilon_i) = \beta_0 + \beta_1 x_i,$$

$$Cov(Y_i, Y_j) = \sigma^2 \qquad i \neq j,$$

y

$$Var(Y_i) = Var(\beta_0 + \beta_1 x_i + \varepsilon_i) = Var(\varepsilon_i) = \sigma^2.$$

El último resultado surge del hecho de que la varianza de una variable aleatoria no varía con respecto a la localización; en este caso, el corrimiento en localización está proporcionado por el término no aleatorio $\beta_0 + \beta_1 x_i$. Por lo tanto, en términos reales, lo que se supone es que para cada calificación promedio x existe una distribución de probabilidad para los salarios iniciales cuya media es una función lineal de x y cuya varianza es la misma para toda x. El modelo proporcionado por (13.1) debe considerarse sólo como una selección inicial para la forma funcional de la curva de regresión. Con base en análisis más apropiados, los cuales se examinarán más adelante, puede ser necesario hacer ajustes y éstos a su vez pueden dar como resultado una ecuación final de predicción diferente de la del modelo inicial.

Para obtener los estimadores de mínimos cuadrados de β_0 y β_1, se generalizará un conjunto de datos consistente en n pares (x_1, y_1), (x_2, y_2), ..., (x_n, y_n), donde los valores de y son las observaciones de la variable aleatoria respuesta. El método de mínimos cuadrados considera la desviación de la observación Y_i de su valor medio y determina los valores de β_0 y β_1 que minimizan la suma de los cuadrados de estas desviaciones. La i-ésima desviación o error es

$$\varepsilon_i = Y_i - (\beta_0 + \beta_1 x_i), \tag{13.2}$$

y la suma de los cuadrados de los errores es

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i)^2. \tag{13.3}$$

Los estimadores de mínimos cuadrados de β_0 y β_1 se obtienen mediante la diferenciación de (13.3) con respecto a β_0 y β_1 y después al igualar cada derivada parcial con cero, es decir

$$\frac{\partial \sum \varepsilon_i^2}{\partial \beta_0} = -2 \sum (Y_i - B_0 - B_1 x_i) = 0,$$

y

$$\frac{\partial \sum \varepsilon_i^2}{\partial \beta_1} = -2 \sum x_i(Y_i - B_0 - B_1 x_i) = 0,$$

donde B_0 y B_1 son los estimadores de mínimos cuadrados* de β_0 y β_1, respectivamente. Al simplificar y distribuir las sumas en estas ecuaciones, se tiene

$$\sum_{i=1}^{n} Y_i = nB_0 + B_1 \sum_{i=1}^{n} x_i$$

y (13.4)

$$\sum_{i=1}^{n} x_i Y_i = B_0 \sum_{i=1}^{n} x_i + B_1 \sum_{i=1}^{n} x_i^2.$$

Las dos ecuaciones dadas por (13.4) se conocen como *ecuaciones normales.*

Dadas las realizaciones $y_1, y_2, \ldots, y_n$, las ecuaciones pueden resolverse para los estimados de mínimos cuadrados b_0 y b_1. Si se dividen ambos mienbros de la primera ecuación entre *n*, se obtiene

$$\frac{\sum y_i}{n} = b_0 + b_1 \frac{\sum x_i}{n};$$

entonces el estimador de mínimos cuadrados de β_0 es

$$b_0 = \frac{\sum_{i=1}^{n} y_i}{n} - b_1 \frac{\sum_{i=1}^{n} x_i}{n} = \bar{y} - b_1 \bar{x}. \tag{13.5}$$

Al sustituir b_0 en la segunda ecuación de (13.4) se obtiene

$$\sum x_i y_i = \left(\frac{\sum y_i}{n} - b_1 \frac{\sum x_i}{n}\right) \sum x_i + b_1 \sum x_i^2,$$

la que, después de resolver para b_1, se reduce a

$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i - \dfrac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}. \tag{13.6}$$

* Muchos autores prefieren designar a los estimadores de mínimos cuadrados con letras cursivas minúsculas. Para mantener la consistencia de la notación con respecto a los capítulos anteriores se designará al estimador de mínimos cuadrados con una letra cursiva en mayúscula y el tipo en minúscula para el estimado *MC*.

Los valores dados por (13.5) y (13.6) son aquellos que minimizan la suma de los cuadrados de los errores.

Dados los estimadores de mínimos cuadrados B_0 y B_1 para la intersección y la pendiente, respectivamente, la recta de regresión estimada para el modelo (13.1) es

$$\hat{Y}_i = B_0 + B_1 x_i \tag{13.7}$$

donde $\hat{Y}_i$ es el estimador para la media de la observación Y_i, la cual corresponde al valor x_i de la variable de predicción. Nótese que si se sustituye (13.5) por B_0 en (13.7) se obtiene una forma alternativa para la recta de regresión estimada, la cual se encuentra dada por

$$\begin{aligned} \hat{Y}_i &= \bar{Y} - B_1\bar{x} + B_1 x_i \\ &= \bar{Y} + B_1(x_i - \bar{x}). \end{aligned} \tag{13.8}$$

Con base en (13.2), la diferencia entre la realización y_i y el valor estimado $\hat{y}_i$ es un estimador del correspondiente error. Este estimador se conoce como el i-ésimo residual y se denota por

$$e_i = y_i - \hat{y}_i. \tag{13.9}$$

De nuevo, nótese que los residuos no son estimados en el sentido clásico de la estimación de parámetros (fijos), sino que son estimadores de los valores de las variables aleatorias no observables ε_i, los cuales se obtienen de la recta de regresión estimada. Los residuos $e_1, e_2, \ldots, e_n$ son muy importantes debido a que proporcionan una abundante información sobre lo que puede faltar del modelo de regresión estimado. Más adelante se darán más detalles con respecto a lo anterior. En este momento se ilustrarán los pesos de cálculo para obtener la recta de regresión estimada para el modelo lineal simple empleando para ello los datos de los salarios. El propósito de esto radica en familiarizar al lector únicamente con el procedimiento de cálculo. De lo contrario, se puede hacer uso de algún paquete estadístico para computadora. Posteriormente, se presentará un listado de computadora para este ejemplo.

En la tabla 13.2, se incluyen los cálculos básicos necesarios para obtener los estimadores de mínimos cuadrados de la intersección y la pendiente. Las últimas cuatro columnas de esta tabla no son necesarias para la determinación de b_0 y b_1. éstas serán empleadas después en otro contexto.

Mediante el empleo de (13.5) y (13.6) el estimador de mínimos cuadrados para la pendiente es

$$b_1 = \frac{830.425 - \dfrac{(45.6)(270.8)}{15}}{139.51 - \dfrac{(45.6)^2}{15}} = 8.12,$$

y el correspondiente estimado de mínimos cuadrados para la intersección es

$$b_0 = \frac{270.8}{15} - (8.12)\frac{45.6}{15} = -6.63.$$

TABLA 13.2 Cálculos básicos para obtener los estimadores de mínimos cuadrados b_0 y b_1 (con base en los datos de salarios dados en la tabla 13.1)

CP x_i	Salario y_i	$x_i y_i$	x_i^2	y_i^2	Salario estimado $\hat{y}_i$	Residuo $y_i - \hat{y}_i$	Cuadrado del residuo $(y_i - \hat{y}_i)^2$
2.95	18.5	54.575	8.7025	342.25	17.32	1.18	1.3924
3.20	20.0	64.000	10.2400	400.00	19.35	0.65	0.4225
3.40	21.1	71.740	11.5600	445.21	20.98	0.12	0.0144
3.60	22.4	80.640	12.9600	501.76	22.60	−0.20	0.0400
3.20	21.2	67.840	10.2400	449.44	19.35	1.85	3.4225
2.85	15.0	42.750	8.1225	225.00	16.51	−1.51	2.2801
3.10	18.0	55.800	9.6100	324.00	18.54	−0.54	0.2916
2.85	18.8	53.580	8.1225	353.44	16.51	2.29	5.2441
3.05	15.7	47.885	9.3025	246.49	18.13	−2.43	5.9049
2.70	14.4	38.880	7.2900	207.36	15.29	−0.89	0.7921
2.75	15.5	42.625	7.5625	240.25	15.70	−0.20	0.0400
3.10	17.2	53.320	9.6100	295.84	18.54	−1.34	1.7956
3.15	19.0	59.850	9.9225	361.00	18.95	0.05	0.0025
2.95	17.2	50.740	8.7025	295.84	17.32	−0.12	0.0144
2.75	16.8	46.200	7.5625	282.24	15.70	1.10	1.2100
Totales 45.6	270.8	830.425	139.5100	4970.12	270.79	0.01	22.8671

De acuerdo con lo anterior, la ecuación estimada de regresión es

$$\hat{y}_i = -6.63 + 8.12\,x_i. \tag{13.10}$$

Al intentar interpretar esta ecuación se tiene que los valores $\hat{y}_i$ son los estimadores para las medias de las distribuciones de probabilidad de los salarios iniciales correspondientes a las calificaciones promedio x_i. Tener una intersección negativa resulta fastidioso, ya que, por ejemplo, si $x = 0.5$, $\hat{y} = -2.57$, lo cual es absurdo. Pero las calificaciones promedio en este conjunto de datos varían de 2.70 a 3.60, por lo tanto, cualquiera que sea la validez que tiene la ecuación estimada de regresión al predecir los salarios iniciales promedio se mantiene, para todos aquellos valores de x que se encuentren entre 2.70 y 3.60. En la práctica, muchas veces se desea predecir la respuesta más allá del intervalo de valores de x para los cuales se obtuvo la ecuación estimada de regresión. Si un valor de x se encuentra muy cercano a este intervalo, la predicción tendrá cierta validez. De otra forma, ésta debe verse con mucho cuidado, ya que la ecuación de regresión estimada puede no ser apropiada para un intervalo de valores más amplio de la variable de predicción.

La interpretación del valor estimado de la pendiente es directa. El incremento estimado en el salario inicial promedio para cada aumento igual a una unidad de la calificación promedio es de 8 120 dólares.

La tercera columna de la derecha en la tabla 13.2, contiene los salarios promedio estimados para cada calificación promedio dada por (13.10). Por ejemplo, si $x = 2.95$, el salario inicial estimado promedio es $\hat{y} = -6.63 + 8.12(2.95) = (13.9)$, miles de dólares. Dado que el correspondiente valor observado es 18.5, de (13.9), $e = 18.5 - 17.32 = 1.18$ es el residuo para $x = 2.95$. En otras palabras, el valor resi-

dual 1.18 es la distancia vertical que existe entre la observación 18.5 y el punto sobre la recta estimada de regresión para $x = 2.95$. Los otros residuos se obtienen de la misma manera y tienen significados similares. La figura 13.2 ilustra los residuos como distancias verticales desde la recta de regresión estimada. Dado que un residuo representa la cantidad en la que un valor estimado falla para predecir la media de la correspondiente observación aleatoria, entre más grandes son las magnitudes de los residuos, mayor tenderá a ser el efecto de la componente aleatoria en el modelo.

Recuérdese que la varianza σ^2 de la variable respuesta es igual a la varianza del error y ésta es constante para todos los valores de la variable de predicción. En general, dado que el valor de σ^2 no se conoce, puede obtenerse un estimador de éste a partir de los estimados de mínimos cuadrados b_0 y b_1. Dado que cada $\hat{y}_i$ estima la media de Y_i, la diferencia $y_i - \hat{y}_i$ representa la desviación de Y_i con respecto a su propia media. La suma de los cuadrados de estas diferencias, dividida entre una constante apropiada, es la forma en la que se determina una varianza. Pero estas diferencias son los residuos; por lo tanto, la suma de los cuadrados de los residuos dividida entre una constante apropiada es un estimador de σ^2. La constante apropiada es $n - 2$, ya que se pierden dos grados de libertad al tener que estimar los dos parámetros β_0 y β_1 antes de obtener $\hat{y}_i$. El estimador de σ^2 se denota como s^2 y está dado por

$$s^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n - 2} = \frac{\sum_{i=1}^{n} e_i^2}{n - 2}. \tag{13.11}$$

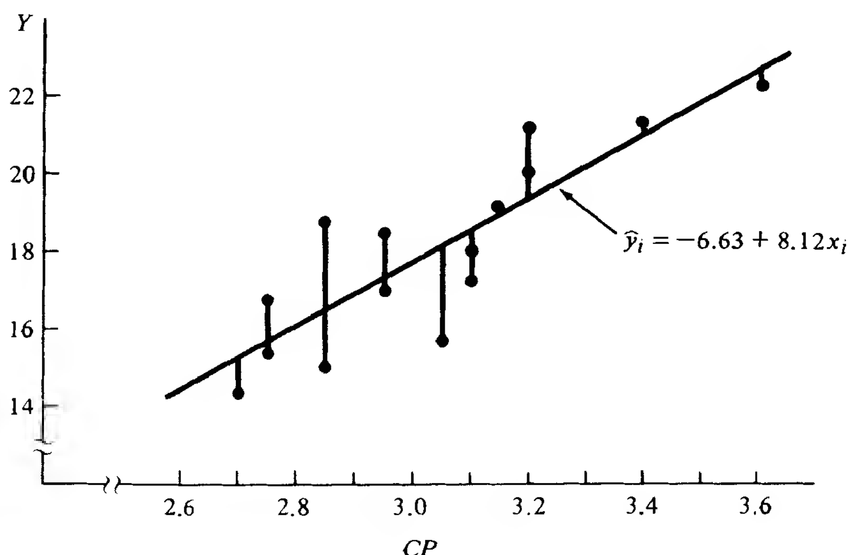


FIGURA 13.2 Residuos como distancias verticales desde la ecuación estimada de regresión

El estimador s^2 recibe el nombre de *varianza residual* o *CME,* y la raíz cuadrada positiva s se conoce como la *desviación estándar residual.* Para el ejemplo de los salarios iniciales, la varianza residual es $s^2 = 22.8671/13 = 1.759$. La varianza residual s^2 es una medida absoluta de qué tan bien se ajusta la recta estimada de regresión a las medias de las observaciones de la variable respuesta. Por lo tanto, en general entre más pequeño sea el valor de s^2, se ajustará mejor al modelo. Puede demostrarse que el estimador S^2 es un estimador no sesgado de σ^2 con tal de que la forma del modelo de regresión sea la correcta. De otra manera, S^2 estima σ^2 más una componente que es el sesgo causado por un error en el modelo.

Cuando se obtiene una recta de regresión por el método de mínimos cuadrados, surgen cierto número de propiedades. Algunas de éstas son las siguientes:

1. $\Sigma_{i=1}^{n} e_i = 0.$
2. $\Sigma_{i=1}^{n} y_i = \Sigma_{i=1}^{n} \hat{y}_i.$
3. $\Sigma_{i=1}^{n} x_i e_i = 0.$

Se demostrará la propiedad 1 y se dejan las correspondientes demostraciones de las propiedades restantes al lector. Debe notarse que la propiedad 2 se obtiene de la primer ecuación dada en (13.14) y la propiedad 3 de la segunda ecuación normal. Para la propiedad 1,

$$\begin{aligned}
\sum_{i=1}^{n} e_i &= \sum_{i=1}^{n} (y_i - \hat{y}_i) \\
&= \sum (y_i - b_0 - b_1 x_i) \\
&= \sum y_i - nb_0 - b_1 \Sigma x_i \\
&= n\bar{y} - n(\bar{y} - b_1\bar{x}) - nb_1\bar{x} \\
&= 0.
\end{aligned}$$

A causa de los errores de redondeo, la suma de los residuos dados en la tabla 13.2 no es exactamente igual a cero. Además, dado que los estimadores *MC* se obtienen mediante la minimización de la suma de los cuadrados de los errores, para este ejemplo el valor mínimo es 22.8671.

13.4 Estimación por máxima verosimilitud para el modelo lineal simple

Puede emplearse el principio de máxima verosimilitud para estimar los parámetros desconocidos en el modelo lineal simple dado por (13.1). Recuérdese que los estimadores de mínimos cuadrados se obtuvieron sin tener que especificar la distribución de probabilidad de los errores aleatorios ε_i. Si se supone que los ε_i son variables aleatorias independientes, normalmente distribuidas, con media cero y varianza σ^2 para toda $i = 1, 2, \ldots, n$, es posible obtener los estimadores de máxima verosimi-

litud de β_0, β_1, y σ^2,es decir, si además de las suposiciones previas se especifica que $\varepsilon_i \sim N(0, \sigma^2)$ para toda $i = 1, 2, \ldots, n$, entonces cada Y_i también se encuentra normalmente distribuida con media $\beta_0 + \beta_1 x_i$ y varianza σ^2, dado que ésta es una función lineal de una variable aleatoria con distribución normal. Los estimadores de máxima verosimilitud se obtienen mediante la maximización de la función de verosimilitud dada por

$$L(y_1, y_2, \ldots, y_n; \beta_0, \beta_1, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{1}{2\sigma^2}(y_1 - \beta_0 - \beta_1 x_1)^2\right]$$

$$\cdots \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{1}{2\sigma^2}(y_n - \beta_0 - \beta_1 x_n)^2\right]$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\right],$$

donde

$$\ln[L(\beta_0, \beta_1, \sigma^2)] = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum(y_i - \beta_0 - \beta_1 x_i)^2.$$

Al tomar las derivadas parciales con respecto a β_0, β_1, y σ^2, y después de igualarlas a cero, puede demostrarse que los estimadores de máxima verosimilitud de β_0 y β_1 son idénticos a los dados por (13.5) y (13.6), respectivamente, y el correspondiente a σ^2 está dado por

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}. \tag{13.12}$$

El estimador de máxima verosimilitud de σ^2 es sesgado pero, para valores grandes de n, la diferencia entre éste y el estimador de mínimos cuadrados no es importante.

El lector puede sorprenderse del por qué la necesidad de tratar con los estimadores de máxima verosimilitud, si éstos son iguales a los estimadores de mínimos cuadrados. Una de estas razones es que los estimadores de máxima verosimilitud tienen propiedades deseables de consistencia, suficiencia y varianza mínima. Además, éstos proporcionan los medios necesarios para el desarrollo de criterios de inferencia para β_0 y β_1.

La suposición de que los errores se encuentran normalmente distribuidos es justificable, debido a que la componente de error en el modelo es, en general, un efecto compuesto que representa muchas perturbaciones pequeñas pero aleatorias, las cuales son independientes de la variable de predicción y se deben a factores que no se encuentran incluidos en el modelo. En todo caso, las desviaciones de la suposición de normalidad para valores grandes de n no son, en general, serias.

13.5 Propiedades generales de los estimadores de mínimos cuadrados

En esta sección se desarrollarán algunas propiedades generales de los estimadores de mínimos cuadrados, por lo que se considerarán algunos criterios que permitan la construcción de intervalos de confianza y la realización de pruebas de hipótesis con respecto a los parámetros de regresión β_0 y β_1. Así mismo, se examinará la estimación de la respuesta media para una x dada y la predicción de una Y en particular para un valor dado de x. En gran medida, el enfoque de esta sección será de carácter teórico.

Considérense los estimadores no sesgados de β_0 y β_1 que son funciones lineales de las observaciones $Y_1, Y_2, \ldots, Y_n$. Si entre todos estos estimadores de β_0 y β_1 existen algunos cuyas varianzas son más pequeñas que las de todos los demás estimadores no sesgados de β_0 y β_1, entonces estos son los *mejores estimadores lineales no sesgados* (*MELI*) de β_0 y β_1. El siguiente teorema conocido generalmente como teorema de Gauss- Markov, garantiza que los estimadores de mínimos cuadrados de β_0 y β_1 son los *MELI* para β_0 y β_1.

Teorema 13.1 Sean las suposiciones para el modelo $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ las mismas que aquellas que se necesitan para la estimación de mínimos cuadrados de β_0 y β_1. Entonces los estimadores de mínimos cuadrados de B_0 y B_1 son los mejores estimadores lineales no sesgados de β_0 y β_1.

Mientras que la demostración del teorema 13.1 se encuentra más allá de los objetivos de este libro, se demostrará que B_0 y B_1 son combinaciones lineales de las observaciones $Y_1, Y_2, \ldots, Y_n$. Lo anterior permitirá demostrar que

$$E(B_1) = \beta_1$$

y

$$Var(B_1) = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \tag{13.13}$$

mientras que

$$E(B_0) = \beta_0$$

y

$$Var(B_0) = \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} (x_i - \bar{x})^2}. \tag{13.14}$$

Para demostrar que B_1 es una combinación lineal de $Y_1, Y_2, \ldots, Y_n$, recuérdese la segunda expresión dada en (13.6). Primero, se desea demostrar que

$$\sum_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y}) = \sum_{i=1}^{n} (x_i - \bar{x}) Y_i.$$

Lo anterior es verdadero, dado que

$$\sum (x_i - \overline{x})(Y_i - \overline{Y}) = \sum (x_i - \overline{x})Y_i - \overline{Y}\sum (x_i - \overline{x});$$

pero $\Sigma(x_i - \overline{x}) = 0$, y

$$\sum (x_i - \overline{x})(Y_i - \overline{Y}) = \sum (x_i - \overline{x})Y_i.$$

De acuerdo con lo anterior,

$$B_1 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})Y_i}{\sum_{i=1}^{n} (x_i - \overline{x})^2},$$

donde las x_i son fijas ya que son valores de una variable de predicción no aleatoria.

Sea

$$c_i = \frac{x_i - \overline{x}}{\sum_{i=1}^{n} (x_i - \overline{x})^2}, \tag{13.15}$$

donde los c_i' son cantidades fijas, dado que las x_i' también son fijas. Entonces el estimador B_1 se expresa como

$$B_1 = \sum_{i=1}^{n} c_i Y_i,$$

la cual es una combinación lineal de las observaciones $Y_1, Y_2, \ldots, Y_n$.

Para demostrar que B_1 es un estimador no sesgado de β_1, se tiene

$$\begin{aligned} E(B_1) &= E\left(\sum_{i=1}^{n} c_i Y_i\right) \\ &= \sum c_i E(Y_i) \\ &= \sum c_i(\beta_0 + \beta_1 x_i) \\ &= \beta_0 \sum_{i=1}^{n} c_i + \beta_1 \sum_{i=1}^{n} c_i x_i. \end{aligned}$$

Pero

$$\sum_{i=1}^{n} c_i = \frac{\sum_{i=1}^{n} (x_i - \overline{x})}{\sum_{i=1}^{n} (x_i - \overline{x})^2} = 0,$$

y

$$\sum_{i=1}^{n} c_i x_i = \frac{\sum_{i=1}^{n} (x_i - \bar{x})x_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = 1.$$

De esta forma,

$$E(B_1) = \beta_1.$$

Dado que por hipótesis las observaciones Y_i no se encuentran correlacionadas por pares, $Cov(Y_i, Y_j) = 0,\ i \neq j$. Entonces, mediante el empleo de la segunda parte del teorema 6.1, se demuestra que la varianza de B_1 está dada por (13.13.) De esta forma se tiene

$$\begin{aligned} Var(B_1) &= Var\left(\sum_{i=1}^{n} c_i Y_i\right) \\ &= \sum c_i^2 Var(Y_i) \\ &= \sum c_i^2 \sigma^2 \\ &= \sigma^2 \sum c_i^2 \\ &= \sigma^2 \sum_{i=1}^{n} \left[\frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}\right]^2 \\ &= \sigma^2 \left\{\sum (x_i - \bar{x})^2 \Big/ \left[\sum (x_i - \bar{x})^2\right]^2\right\} \\ &= \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}. \end{aligned}$$

La raíz cuadrada de $Var(B_1)$ es la desviación estándar* del estimador de mínimos cuadrados de la pendiente y está dada por

$$d.e.(B_1) = \frac{\sigma}{\left[\sum (x_i - \bar{x})^2\right]^{1/2}}.$$

Dado que, en general, la desviación estándar σ del error es desconocida, puede obtenerse un estimador de $d.e.(B_1)$ al reemplazar σ por la desviación estándar residual s, como está dada por (13.11). De esta forma, un estimado de la desviación estándar de B_1 es

* También conocida como error estándar.

$$s(B_1)^* = \frac{s}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{1/2}}. \qquad (13.16)$$

Ahora considérese el estimador de mínimos cuadrados de la intersección desconocida β_0. Dado que el estimador de mínimos cuadrados es

$$B_0 = \bar{Y} - B_1\bar{x},$$

y ya que el estimador de mínimos cuadrados de la pendiente es una combinación lineal de las observaciones $Y_1, Y_2, \ldots, Y_n$, entonces también B_0 es una combinación lineal de las observaciones. Para demostrar que B_0 es un estimador no sesgado de β_0, se tiene

$$\begin{aligned} E(B_0) &= E(\bar{Y} - B_1\bar{x}) \\ &= \frac{\sum_{i=1}^{n} E(Y_i)}{n} - \bar{x}\,E(B_1) \\ &= \frac{\sum(\beta_0 + \beta_1 x_i)}{n} - \beta_1\bar{x} \\ &= \frac{n\beta_0 + \beta_1\sum x_i}{n} - \beta_1\bar{x} \\ &= \beta_0 + \beta_1\bar{x} - \beta_1\bar{x} \\ &= \beta_0. \end{aligned}$$

Para demostrar que $Var(B_0)$ está dada por (13.14), de nuevo se empleará la segunda parte del teorema 6.1 y el hecho de que B_0 y B_1 son combinaciones lineales de variables aleatorias no correlacionadas. Dado que $B_0 = \bar{Y} - B_1\bar{x}$,

$$\begin{aligned} Var(B_0) &= Var(\bar{Y} - B_1\bar{x}) \\ &= Var\left(\frac{\sum Y_i}{n} - \bar{x}\sum c_iY_i\right) \\ &= Var\left[\sum_{i=1}^{n}\left(\frac{Y_i}{n} - \bar{x}c_iY_i\right)\right] \\ &= Var\left[\sum\left(\frac{1}{n} - \bar{x}c_i\right)Y_i\right] \end{aligned}$$

* Se empleará la notación más conveniente $s^2(T)$ y $s(T)$ para denotar, respectivamente, la varianza y la desviación estándar estimadas de un estimador. T.

$$= \sum \left(\frac{1}{n} - \bar{x} c_i \right)^2 Var(Y_i)$$

$$= \sigma^2 \sum \left(\frac{1}{n^2} - \frac{2\bar{x}c_i}{n} + \bar{x}^2 c_i^2 \right)$$

$$= \sigma^2 \left(\frac{1}{n} - \frac{2\bar{x}}{n} \sum c_i + \bar{x}^2 \sum c_i^2 \right).$$

Al sustituir (13.15) por c_i y al recordar que $\Sigma c_i = 0$, se tiene

$$Var(B_0) = \sigma^2 \left\{ \frac{1}{n} + \bar{x}^2 \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\left[\sum_{i=1}^{n} (x_i - \bar{x})^2 \right]^2} \right\}$$

$$= \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} \right]$$

Finalmente, si se sustituye $\bar{x}^2 = (\Sigma x_i)^2/n^2$, se obtiene

$$Var(B_0) = \sigma^2 \left[\frac{1}{n} + \frac{\left(\sum x_i \right)^2}{n^2 \sum (x_i - \bar{x})^2} \right]$$

$$= \sigma^2 \left[\frac{n \sum (x_i - \bar{x})^2 + \left(\sum x_i \right)^2}{n^2 \sum (x_i - \bar{x})^2} \right]$$

$$= \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} (x_i - \bar{x})^2}.$$

Entonces, un estimador de la desviación estándar de B_0 es

$$s(B_0) = s \left[\frac{\sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} (x_i - \bar{x})^2} \right]^{1/2} \qquad (13.17)$$

Es interesante notar que las varianzas de B_0 y B_1 son funciones de los valores x_i para los cuales se observa la variable respuesta. En particular, para el estimador de la pendiente B_1, $Var(B_1)$ tiene un valor máximo cuando $\Sigma(x_i - \bar{x})^2$ tiene un valor máximo. Pero $\Sigma(x_i - \bar{x})^2$ es máxima cuando la distancia entre los valores de x_i es la más grande. Esto ocurre cuando se escoge observar la respuesta sólo en los valores de los extremos del intervalo de variación de la variable de predicción, es decir, si verdaderamente el modelo de regresión es lineal, entonces deberán tomarse $n/2$ observaciones en un extremo y $n/2$ en el otro para obtener la mejor eficiencia posible al estimar la pendiente de la línea recta. Lo anterior es lógico ya que sólo se necesitan dos puntos para definir una línea recta; sin embargo, en la práctica, no es muy común el hecho de saber que la función de regresión es lineal de manera tal, que no sería prudente seleccionar los extremos del intervalo de x como puntos de observación y minimizar la varianza del estimador de la pendiente. Una alternativa más segura consiste en tener puntos de observación espaciados de igual forma sobre todo el intervalo de interés de la variable de predicción.

Para el modelo lineal simple, la recta de regresión estimada dada por (13.7) permite obtener un estimador para la media de la variable de respuesta para un valor específico de la variable de predicción. Sea x_p este valor en particular y para el cual se desea estimar la media de la variable respuesta Y_p. Entonces el estimador es $\hat{y}_p = b_0 + b_1 x_p$. Para el mismo conjunto de valores de x existe una variación muestra a muestra en el estimador $\hat{Y}_p$, dado que existe una variación del mismo tipo para los estimadores de mínimos cuadrados B_0 y B_1. Puede observarse que lo anterior es cierto para el ejemplo del salario inicial, ya que no se espera tener la misma recta de regresión estimada si se selecciona otro conjunto de estudiantes con las mismas *CP* que los primeros.

Considérese que la determinación de la media, y la varianza de $\hat{Y}_p$. $\hat{Y}_p$ es un estimador no sesgado de la media de Y_p, dado que

$$E(\hat{Y}_p) = E(B_0 + B_1 x_p) = \beta_0 + \beta_1 x_p = E(Y_p).$$

Para obtener la varianza de $\hat{Y}_p$, se hará uso de la misma técnica empleada para la varianza de B_0. De (13.8) se tiene

$$\begin{aligned}
Var(\hat{Y}_p) &= Var[\bar{Y} + B_1(x_p - \bar{x})] \\
&= Var\left[\frac{\sum Y_i}{n} + (x_p - \bar{x}) \sum c_i Y_i\right] \\
&= Var\left\{\sum_{i=1}^{n} \left[\frac{1}{n} + c_i(x_p - \bar{x})\right] Y_i\right\} \\
&= \sum \left[\frac{1}{n} + c_i(x_p - \bar{x})\right]^2 Var(Y_i) \\
&= \sigma^2\left[\frac{1}{n} + \frac{2(x_p - \bar{x})}{n} \sum c_i + (x_p - \bar{x})^2 \sum c_i^2\right] \\
&= \sigma^2 \left[\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right] \qquad (13.18)
\end{aligned}$$

Por lo tanto, un estimador de la desviación estándar de $\hat{Y}_p$ está dado por

$$s(\hat{Y}_p) = s\left[\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right]^{1/2} \tag{13.19}$$

Supóngase que en lugar de estimar la media de Y_p en x_p, se desea predecir un valor particular de Y_p que se observaría si se impusiera un valor x_p para la variable de predicción. Por ejemplo, dada la ecuación de regresión estimada, ¿cuál podría ser el valor del salario inicial para un estudiante en particular con un *CP* conocido? Aunque se trata de un solo estudiante, puede ser razonable predecir el salario inicial promedio para un *CP* dado. De esta forma, si se desea estimar la media de Y_p o un valor particular de Y_p para x_p, el valor estimado es el mismo y está dado por (13.7). Pero es evidente que la varianza de la predicción para este último caso puede tener un valor más grande, ya que ésta no sólo considera la variación muestra a muestra de $\hat{Y}_p$, sino también la variación inherente de la distribución de probabilidad de Y_p. Si se supone que el valor predicho de Y_p para x_p es independiente de la muestra que proporciona la recta de regresión estimada, la covarianza de Y_p y $\hat{Y}_p$ es cero. Entonces

$$\begin{aligned} Var(\hat{Y}_{\text{part}}) &= Var(Y_p) + Var(\hat{Y}_p) \\ &= \sigma^2 + \sigma^2\left[\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right] \\ &= \sigma^2\left[1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right], \end{aligned} \tag{13.20}$$

donde $\hat{Y}_{\text{part}}$ denota la predicción particular para Y_p en x_p. Del análisis anterior se obtiene que un estimador de la desviación estándar de $\hat{Y}_{\text{part}}$ está dado por

$$s(\hat{Y}_{\text{part}}) = s\left[1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right]^{1/2} \tag{13.21}$$

Mediante el empleo de los datos que aparecen en la tabla 13.2, se ilustra el cálculo de las varianzas y las desviaciones estándar de los estimadores de mínimos cuadrados B_1 y B_0. Dado que

$$\sum (x_i - \bar{x})^2 = \sum x_i^2 - \left(\sum x_i\right)^2/n = 0.886$$

y $s^2 = 1.759$,

$$s^2(B_1) = \frac{1.759}{0.886} = 1.985,$$

y

$$s(B_1) = 1.409.$$

De manera similar,

$$s^2(B_0) = \frac{(1.759)(139.51)}{(15)(0.886)} = 18.465$$

y

$$s(B_0) = 4.297.$$

Si se continúa con este ejemplo, supóngase que se desea estimar la media de la distribución de salarios inciales cuando el *CP* es $x_p = 3.25$. Nótese que este valor no es uno de los valores de x que dieron origen a la recta de regresión estimada, pero se encuentra dentro del intervalo definido por estos valores. De (13.10) y (13.18) la media y la varianza estimadas para $x_p = 3.25$, son

$$\hat{y}_p = -6.63 \pm 8.12(3.25) = 19.76$$

y

$$s^2(\hat{Y}_p) = 1.759\left[\frac{1}{15} + \frac{(3.25 - 3.04)^2}{0.886}\right] = 0.205,$$

respectivamente. De esta forma, la desviación estándar estimada es $\sqrt{0.205} = 0.453$ miles de dólares. Si se desea predecir el salario inicial real para un estudiante en particular con una *CP* de 3.25, el valor estimado sería aún de 19.76 miles de dólares, pero la varianza estimada sería de

$$1.759\left[1 + \frac{1}{15} + \frac{(3.25 - 3.04)^2}{0.886}\right] = 1.964,$$

o una desviación estándar de 1.401 miles de dólares.

En esta sección se han determinado las medias y las varianzas de los estimadores B_0, B_1, $\hat{Y}_p$ y $\hat{Y}_{\text{part}}$, pero aún no se han desarrollado sus distribuciones de muestreo. Para realizar esto es necesario suponer el caso de la teoría normal de la sección anterior, en el que se supone que cada error aleatorio ε_i tiene una distribución normal con media cero y varianza σ^2 para toda $i = 1, 2, \ldots, n$. Por lo tanto, las observaciones $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias independientes y distribuidas en forma normal con medias $\beta_0 + \beta_1 x_i$ y varianza común σ^2, para $i = 1, 2, \ldots, n$.

Para obtener la distribución de la muestra para el estimador de la pendiente B_1, bajo el caso de la teoría normal, sólo necesita recordarse que B_1 es una combinación lineal de variables aleatorias normalmente distribuidas y, de esta forma, la combinación es una variable aleatoria con distribución normal, media β_1 y varianza dadas por (13.13). Al recordar la definición de una variable aleatoria t de Student puede demostrarse que la distribución de la cantidad

$$(B_1 - \beta_1)/s(B_1)$$

es la t de Student con $n - 2$ grados de libertad. El estimador B_0 también es una combinación lineal de variables aleatorias normalmente distribuidas. Así, B_0 también es normalmente distribuida, con media β_0 y varianza dadas por (13.14). Además, se puede mostrar que la cantidad

$$(B_0 - \beta_0)/s(B_0)$$

es una variable aleatoria de la t de Student con $n - 2$ grados de libertad. Como se verá en la siguiente sección, estos resultados permiten la formulación de inferencias estadísticas con respecto a los parámetros desconocidos β_0 y β_1.

Bajo el caso de la teoría normal, el estimador $\hat{Y}_p = B_0 + B_1 x_p$ de la media de Y_p para x_p también se encuentra normalmente distribuido con media $E(Y_p)$ y varianza dada por (13.18), ya que ésta es una combinación lineal de variables aleatorias normalmente distribuidas. Entonces, la distribución de

$$[\hat{Y}_p - E(Y_p)]/s(\hat{Y}_p)$$

es la t de Student con, de nuevo, $n - 2$ grados de libertad. También se obtiene un resultado similar para la predicción $\hat{Y}_{\text{part}}$ para una respuesta en particular Y_p correspondiente a x_p. Así, resulta comprensible el porqué $n - 2$ grados de libertad, ya que la determinación de la recta de regresión necesita la estimación de los dos parámetros de regresión β_0 y β_1.

13.6 Inferencia estadística para el modelo lineal simple

En la sección precedente se examinaron las propiedades teóricas de los estimadores para el modelo lineal simple. En esta sección se emplearán esas propiedades para llevar a cabo un analisis de regresión, es decir, se desarrollarán pruebas de hipótesis e intervalos de confianza para las cantidades de interés en este modelo.

El parámetro clave del modelo lineal simple

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

tiene que ser la pendiente β_1. Si la respuesta Y se encuentra relacionada en forma lineal con la variable de predicción x, la pendiente β_1 tiene que ser diferente de cero. De otra forma, no existe ninguna relación lineal entre Y y x. Un procedimiento inferencial natural para β_1 es construir un intervalo de confianza del $100(1 - \alpha)\%$ para β_1. Si este intervalo no contiene el valor cero, entonces es razonable concluir que β_1 es diferente de cero y que Y y x están, en algún grado, relacionados en forma lineal.

Recuérdese que bajo el caso de la teoría normal, la variable aleatoria $(B_1 - \beta_1)/s(B_1)$ tiene una distribución t de Student con $n - 2$ grados de libertad. Entonces

$$P[B_1 - t_{1-\alpha/2,\, n-2}\, s(B_1) < \beta_1 < B_1 + t_{1-\alpha/2,\, n-2}\, s(B_1)] = 1 - \alpha,$$

o la probabilidad de que el intervalo aleatorio $[B_1 - t_{1-\alpha/2,\, n-2}\, s(B_1),\ B_1 + t_{1-\alpha/2,\, n-2}\, s(B_1)]$ contenga el valor real de la pendiente β_1 es $1 - \alpha$. Al reemplazar el estimador de mínimos cuadrados B_1 por su estimador dado por (13.6), el intervalo de confianza del $100(1 - \alpha)\%$ para β_1 es

$$b_1 \pm t_{1-\alpha/2,\, n-2}\, s(B_1),$$

donde la desviación estándar estimada $s(B_1)$ está dada por (13.16). Como ejemplo, recuérdese la recta de regresión estimada $\hat{y}_i = -6.63 + 8.12x_i$ para los datos de

los salarios iniciales. Dado que $b_1 = 8.12$ y $s(B_1) = 1.409$, entonces un intervalo del 95% de confianza para β_1 es

$$8.12 \pm (2.160)(1.409) = (5.08, 11.16),$$

donde $t_{0.975,\,13} = 2.160$. La interpretación de este intervalo es la siguiente: supóngase que se toman muestras repetidas, cada una del mismo tamaño n, de la variable de respuesta para algunos de los valores de x que producen la recta estimada $\hat{y}_i = -6.63 + 8.12x_i$, construyéndose para cada muestra un intervalo de confianza del 95% para β_1. Por lo tanto, el 95% de todos estos intervalos incluirá el valor real de la pendiente β_1.

Considérese la prueba de la hipótesis nula

$$H_0: \beta_1 = \beta_{1_0}$$

contra la alternativa

$$H_1: \beta_1 \neq \beta_{1_0}$$

donde β_{1_0} es el valor propuesto de la pendiente desconocida β_1. Bajo H_0, la estadística

$$T = \frac{B_1 - \beta_{1_0}}{s(B_1)}$$

tiene una distribución t de Student con $n - 2$ grados de libertad. De esta forma, para un tamaño dado del error de tipo I puede tomarse una decisión, en forma fácil, con base en la evidencia de la muestra. Nótese que también es posible tener hipótesis alternativas unilaterales.

Al igual que en los casos ya analizados, cualquier valor propuesto de β_1 que se encuentre en el correspondiente intervalo de confianza, causará una equivocación al rechazar a H_0. En general, el valor propuesto es el cero; es decir, la hipótesis nula establece que no existe ninguna asociación lineal entre x y Y, así que el valor de la estadística de prueba es

$$t = b_1/s(B_1).$$

Como ejemplo, considérese la prueba de la hipótesis nula

$$H_0: \beta_1 = 0$$

contra la alternativa

$$H_1: \beta_1 > 0$$

para el ejemplo de los salarios iniciales contra *CP*. Se ha seleccionado una hipótesis alternativa unilateral, ya que el sentido común dicta que si existe una relación lineal entre *CP* y el salario inicial, la pendiente deberá ser positiva. Para $\alpha = 0.01$; entonces $t_{0.99,\,13} = 2.650$, y

$$t = 8.12/1.409 = 5.76.$$

Por lo tanto, se rechaza la hipótesis nula de que la pendiente es cero. Este resultado, junto con el intervalo de confianza para β_1, sugiere que el salario inicial promedio se encuentra influenciado, en forma lineal, por la calificación promedio.

También puede construirse un intervalo de confianza para el parámetro de intersección β_0 en forma similar. Dado que $(B_0 - \beta_0)/s(B_0) \sim t$ de Student con $n-2$ grados de libertad,

$$P[B_0 - t_{1-\alpha/2,\ n-2}s(B_0) < \beta_0 < B_0 + t_{1-\alpha/2,\ n-2}s(B_0)] = 1 - \alpha$$

es un intervalo aleatorio para β_0 con probabilidad $1 - \alpha$. Por lo tanto, un intervalo de confianza del $100(1 - \alpha)\%$ para β_0 es

$$b_0 \pm t_{1-\alpha/2,\ n-2}s(B_0)$$

donde b_0 es el estimador de mínimos cuadrados y $s(B_0)$ es la desviación estándar estimada. De nuevo, para el ejemplo de los salarios iniciales, un intervalo de confianza del 99% para β_0 es

$$-6.63 \pm (3.012)(4.297) = (-19.57, 6.31).$$

El lector debe darse cuenta que el significado de un intervalo como el anterior no es del todo aparente, ya que el modelo de regresión no tiene sentido si la calificación promedio es cero. En general, deben evitarse las inferencias con respecto a la intersección, a menos que exista un valor de la respuesta para $x = 0$.

Ahora, considérese la estimación por intervalo de la media de Y_p para x_p. Recuérdese que bajo el caso de la teoría normal, el estimador $\hat{Y}_p = B_0 + B_1 x_p$ tiene una distribución normal con media $E(Y_p)$ y varianza dada por (13.18) y la distribución de muestreo de $[\hat{Y}_p - E(Y_p)]/s(\hat{Y}_p)$ es la t de Student con $n-2$ grados de libertad. Entonces la probabilidad del intervalo aleatorio

$$\hat{Y}_p - t_{1-\alpha/2,\ n-2}s(\hat{Y}_p) < E(Y_p) < \hat{Y}_p + t_{1-\alpha/2,\ n-2}s(\hat{Y}_p)$$

es $1-\alpha$, y un intervalo de confianza del $100(1 - \alpha)\%$ para $E(Y_p)$ es

$$\hat{y}_p \pm t_{1-\alpha/2,\ n-2}s(\hat{Y}_p).$$

Para el ejemplo de los salarios iniciales, supóngase que se desea construir un intervalo de confianza del 95% para la media de Y_p en $x_p = 2.80$. El valor estimado es

$$\hat{y}_p = -6.63 + 8.12(2.80) = 16.11$$

y la desviación estándar estimada es

$$s(\hat{Y}_p) = \left\{1.759\left[\frac{1}{15} + \frac{(2.80 - 3.04)^2}{0.886}\right]\right\}^{1/2} = 0.481.$$

Dado que $t_{0.975,\ 13} = 2.160$, un intervalo de confianza del 95% para $E(Y_p)$ es

$$16.11 \pm (2.160)(0.481) = (15.07, 17.15).$$

Al seguir este procedimiento, pueden obtenerse intervalos de confianza del 95% para $E(Y_p)$ para distintos valores de la variable de predicción. Los resultados se encuentran resumidos en la tabla 13.3.

TABLA 13.3 Intervalos de confianza para los salarios iniciales medios

x_p	$\hat{y}_p$	$s(\hat{Y}_p)$	*Intervalo de confianza del 95%*
2.60	14.48	0.708	(12.95, 16.01)
2.70	15.29	0.589	(14.02, 16.56)
2.80	16.11	0.481	(15.07, 17.15)
2.90	16.92	0.395	(16.07, 17.77)
3.00	17.73	0.347	(16.98, 18.48)
3.10	18.54	0.353	(17.78, 19.30)
3.20	19.35	0.410	(18.46, 20.24)
3.30	20.17	0.501	(19.09, 21.25)
3.40	20.98	0.612	(19.66, 22.30)
3.50	21.79	0.733	(20.21, 23.37)
3.60	22.60	0.860	(20.74, 24.46)

Para ilustrar la natu.aleza de estos intervalos de confianza, cuando se comparan con los valores de la variable de predicción, se grafica la recta estimada de regresión y después los límites inferior y superior de cada intervalo contra x_p. El resultado se ilustra en la figura 13.3. Nótese que los límites inferior y superior forman dos hipérbolas con respecto a la recta de regresión estimada. La distancia vertical entre cada

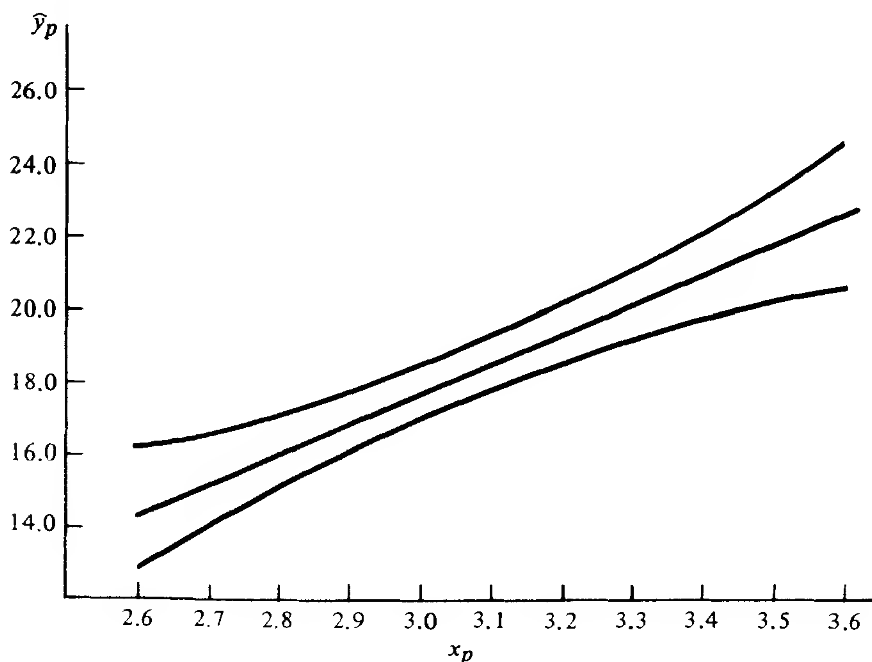


FIGURA 13.3 Intervalos de confianza y la recta de regresión estimada

curva y la recta de regresión es más pequeña para el punto $\bar{x} = 3.04$ y aumenta, en forma simétrica, en ambas direcciones al alejarse de $\bar{x}$. Si se plantea en forma sencilla, los resultados anteriores indican que la predicción de $E(Y_p)$ es más confiable (varianza más pequeña) alrededor de la mitad de los valores de x obtenidos por medio de la ecuación de regresión que en los extremos del intervalo de valores x.

Recuérdese que el usuario puede estar más interesado en predecir una respuesta particular para una x dada, que en estimar la respuesta media para ese mismo valor x. Mientras que el valor predicho puede ser el mismo en cualquier caso, la variabilidad del estimado con respecto a la respuesta en particular será decididamente más grande que la correspondiente a la respuesta media. Dado que, bajo el caso de la teoría normal, la cantidad $[\hat{Y}_{\text{part}} - Y_p]/s(\hat{Y}_{\text{part}})$ es una variable aleatoria t de Student con $n - 2$ grados de libertad entonces, para un α dado,

$$P[\hat{Y}_{\text{part}} - t_{1-\alpha/2,\, n-2}\, s(\hat{Y}_{\text{part}}) < Y_p < \hat{Y}_{\text{part}} + t_{1-\alpha/2,\, n-2}\, s(\hat{Y}_{\text{part}})] = 1 - \alpha.$$

Con base en este resultado puede obtenerse lo que, en general, recibe el nombre de *intervalo de predicción* para la observación Y_p. Un intervalo de predicción es el análogo del intervalo de confianza. Un intervalo de predicción del $100(1 - \alpha)\%$ para una observación particular Y_p, es

$$\hat{y}_{\text{part}} \pm t_{1-\alpha/2,\, n-2}\, s(\hat{Y}_{\text{part}}).$$

Como ejemplo, se construirá un intervalo de predicción del 95% para el salario inicial de un recién graduado con una calificación promedio de 2.80. El valor predicho puede ser el mismo que el de la respuesta media,

$$\hat{y}_{\text{part}} = -6.63 + 8.12(2.80) = 16.11;$$

pero la desviación estándar estimada es

$$s(\hat{Y}_{\text{part}}) = \left\{1.759\left[1 + \frac{1}{15} + \frac{(2.80 - 3.04)^2}{0.886}\right]\right\}^{1/2} = 1.411,$$

la cual es mucho más grande que el valor comparable de 0.481 para $\hat{Y}_p$. Por lo tanto, un intervalo de predicción del 95% para Y_p, es

$$16.11 \pm (2.160)(1.411) = (13.06, 19.16).$$

En la tabla 13.4 se proporcionan los intervalos de predicción del 95% para las observaciones de la respuesta correspondiente a cada uno de los valores de x que se encuentran en la tabla 13.3 y que no son parte del conjunto original que dio origen a la ecuación de regresión estimada. Como era de esperarse, los intervalos de predicción para las observaciones individuales de la respuesta son mucho más grandes que los correspondientes intervalos de confianza para la media de la misma.

Ya que el análisis de regresión se basa en la teoría normal, es apropiado formular un comentario con respecto a las consecuencias sobre la inferencia cuando las distribuciones de probabilidad de los errores aleatorios no son normales. Si la desviación con respecto a la normalidad no es muy grande, las distribuciones de muestreo de los

TABLA 13.4 Intervalos de predicción para los salarios iniciales individuales

x_p	$\hat{y}_{part}$	$s(\hat{Y}_{part})$	*Intervalo de predicción del 95%*
2.60	14.48	1.503	(11.23, 17.73)
2.80	16.11	1.411	(13.06, 19.16)
2.90	16.92	1.384	(13.93, 19.91)
3.00	17.73	1.371	(14.77, 20.69)
3.30	20.17	1.418	(17.11, 23.23)
3.50	21.79	1.515	(18.52, 25.06)

estimadores serán muy cercanas a la normalidad y se acercarán a ésta conforme aumente el tamaño de la muestra. Bajo estas condiciones, la distribución t de Student sigue siendo muy robusta y proporciona aproximaciones muy cercanas a los niveles de confianza propuestos.

13.7 El uso del análisis de varianza

El análisis de regresión para el modelo lineal sencillo también abarca la aplicación de la técnica del análisis de varianza analizada en el capítulo 12. En síntesis, la técnica del análisis de varianza proporciona sólo un medio alternativo al de la sección 13.6 para probar la hipótesis nula de que la pendiente es cero. Sin embargo, permite una comprensión natural del problema y por lo tanto es muy útil para el análisis de modelos más complicados, lo cual se hará más adelante.

Recuérdese que la técnica del análisis de varianza divide la variación total de las observaciones en sus partes componentes de acuerdo con el modelo propuesto. En esencia, para el modelo lineal simple la variación total es la suma de dos componentes: la causada por el término no aleatorio $\beta_1 x$, y la que se debe al error aleatorio ε. Dado que lo que se pretende es que la recta estimada de regresión explique la mayor cantidad posible de la variación total, la contribución del término $\beta_1 x$ debe ser sustancial. El resultado anterior implicaría que las variables respuesta y predicción están relacionadas en forma lineal. Si $\beta_1 = 0$, no existe una asociación lineal entre x y Y.

Para desarrollar el enfoque del análisis de varianza, se seguirá el procedimiento establecido en el capítulo 12. Considérese la desviación de la observación Y_i de la media de las observaciones $\overline{Y}$. Por el momento, supóngase que todas las observaciones Y_i son iguales entre sí, así que la pendiente β_1 debe ser cero, $\varepsilon_i = 0$, y $Y_i = \overline{Y}$ para toda i. Por otro lado, si la magnitud de la desviación $Y_i - \overline{Y}$ es mayor que cero, ésta deberá atribuirse a las componentes del modelo.

Para la desviación

$$Y_i - \overline{Y}$$

supóngase que se suma y se resta el estimador $\hat{Y}_i$ para la media de Y_i, tal como se obtiene de la ecuación de regresión. Entonces

$$\begin{aligned} Y_i - \overline{Y} &= Y_i - \overline{Y} + \hat{Y}_i - \hat{Y}_i \\ &= \hat{Y}_i - \overline{Y} + Y_i - \hat{Y}_i. \end{aligned}$$

De esta forma, la desviación total de la observación Y_i con respecto a la media $\overline{Y}$, es la suma de la desviación de la respuesta media estimada $\hat{Y}_i$ de $\overline{Y}$ y la desviación de Y_i con respecto a $\hat{Y}_i$. Nótese que la última diferencia es el estimador para el *i*-ésimo residuo, el cual representa la distancia vertical desde la respuesta observada al punto correspondiente sobre la recta de regresión estimada. Las desviaciones $Y_i - \hat{Y}_i$ representan la contribución a la componente de error a la variación total. Recuérdese que $\hat{Y}_i$ estima la media de Y_i para x_i. Si la variable de predicción no tiene ningún efecto lineal sobre la respuesta, entonces $\hat{Y}_i$ es virtualmente igual a $\overline{Y}$ para toda *i;* es decir, $\beta_1 = 0$, y el estimador de mínimos cuadrados de β_0 es $\overline{Y}$. Si la magnitud de la desviación de $\hat{Y}_i - \overline{Y}$ es grande, entonces se tiene un efecto lineal de x sobre Y ($\beta_1 \neq 0$).

Para proseguir con el enfoque del análisis de varianza se tomará el cuadrado de ambos miembros de la identidad

$$Y_i - \overline{Y} = \hat{Y}_i - \overline{Y} + Y_i - \hat{Y}_i,$$

y se sumarán para todas las observaciones $\Sigma(Y_i - \overline{Y})^2$. Entonces se tiene

$$\sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 + \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 + 2\sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})(Y_i - \hat{Y}_i).$$

Para demostrar que los productos cruzados son cero, se vuelve a escribir la última suma como

$$\begin{aligned}\sum (\hat{Y}_i - \overline{Y})(Y_i - \hat{Y}_i) &= \sum [\hat{Y}_i(Y_i - \hat{Y}_i) - \overline{Y}(Y_i - \hat{Y}_i)] \\ &= \sum \hat{Y}_i(Y_i - \hat{Y}_i) - \overline{Y}\sum (Y_i - \hat{Y}_i).\end{aligned}$$

De acuerdo con la propiedad 1 de la recta de regresión estimada, examinada en la sección 13.3, la segunda suma es cero. La primera suma puede escribirse como

$$\begin{aligned}\sum \hat{Y}_i(Y_i - \hat{Y}_i) &= \sum \hat{Y}_i\, e_i \\ &= \sum (B_0 + B_1x_i)\, e_i \\ &= B_0\sum e_i + B_1\sum x_i\, e_i \\ &= 0,\end{aligned}$$

Dado que Σe_i y $\Sigma x_i\, e_i$ son cero de las propiedades 1 y 3, respectivamente, por lo tanto la ecuación fundamental del análisis de regresión es

$$\sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 + \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2. \qquad (13.22)$$

De acuerdo con la terminología dada en el capítulo 12, el término $\Sigma(Y_i - \overline{Y})^2$ es la suma total de cuadrados *STC* la cual toma en cuenta la variación total de las observaciones Y_i con respecto a su media sin considerar la variable de predicción. Las compo-

nentes de *STC* son la suma de los cuadrados de los errores $SCE = \Sigma(Y_i - \hat{Y}_i)^2$ y la *suma de los cuadrados de la regresión* $SCR = \Sigma(\hat{Y}_i - \bar{Y})^2$. *SCE* toma en cuenta la variación de las observaciones con respecto a la recta de regresión estimada. Si todas las observaciones se encuentran sobre la recta estimada, el valor de todos los residuos es cero y $SCE = 0$. Se desprende el hecho de que entre más grande es el valor de *SCE*, mayor es la contribución de la componente de error a la variación de las observaciones, o mayor es la incertidumbre cuando se estima la respuesta mediante el uso de la ecuación de regresión. Por otro lado, *SCR* representa la variación de la observación que es atribuible al efecto lineal de x sobre Y. Si la pendiente de la recta estimada de regresión es cero, entonces $SCR = 0$. De esta forma, entre más grande es la proporción de *SCR* con respecto a *SCT*, mayor será la cantidad de la variación en las observaciones que puede explicarse mediante el término lineal $\beta_1 x$

¿Cuál es el número de grados de libertad asociado con cada término de (13.22)? Recuérdese la definición del número de grados de libertad asociados con una suma de cuadrados dada en el capítulo 12. Para *STC* existen $n - 1$ grados de libertad ya que se pierde uno por causa de la restricción lineal $\Sigma(Y_i - \bar{Y}) = 0$ entre las observaciones Y_i. Nótese que *SCE* es el numerador de la expresión (13.11) para el cálculo de la varianza residual, así que el número de grados de libertad para *SCE* será de $n - 2$.* Dado que los grados de libertad son aditivos,

$$\text{gl(SCR)} = \text{gl(STC)} - \text{gl(SCE)},$$

y *SCR* tiene un grado de libertad. Como se observará posteriormente, cuando se traten modelos más complicados, el número de grados de libertad para *SCR* será siempre igual al número de parámetros de regresión en el modelo, sin contar a β_0.

Para el análisis de varianza se buscará una estadística para probar la hipótesis nula

$$H_0: \beta_1 = 0$$

contra la alternativa

$$H_1: \beta_1 \neq 0.$$

En general, H_0 se conoce como la hipótesis de regresión no lineal entre x y Y. Si se supone el caso de la teoría normal, entonces bajo la hipótesis nula las observaciones Y_i son n variables aleatorias independientes normalmente distribuidas con la misma media $\mu = \beta_0$ y varianza σ^2. Por lo tanto, puede demostrarse que SCR/σ^2 y SCE $/\sigma^2$ son dos variables aleatorias independientes con una distribución chi-cuadrada con 1 y $n - 2$ grados de libertad, respectivamente. Entonces, del teorema 7.8, la variable aleatoria

$$F = \frac{\dfrac{\text{SCR}/\sigma^2}{1}}{\dfrac{\text{SCE}/\sigma^2}{n-2}} = \frac{\text{SCR}/1}{\text{SCE}/(n-2)} = \text{CMR}/\text{CME} \tag{13.23}$$

* Los dos grados de libertad que se pierden se deben a las dos restricciones lineales dadas por las propiedades 1 y 3 de la sección 13.3.

tiene una distribución F con 1 y $n-2$ grados de libertad, donde el cuadrado medio del error es igual a la varianza residual.

Para llegar a la región de rechazo apropiada para H_0 sugerida por (13.23), se empleará la intuición. Con base en un conjunto dado de datos, un valor grande de *CME* comparado con *CMR* implicará ajuste pobre y sugerirá la ausencia de una asociación lineal entre x y Y. Pero un valor relativamente pequeño para *CME* implicará el hecho de que una porción considerable de la variación en las observaciones es atribuible a un efecto lineal de x sobre Y. Por lo tanto, la hipótesis nula de regresión no lineal entre x y Y debe rechazarse siempre que el valor de (13.23) sea relativamente grande. De otro modo, la evidencia experimental no apoya el rechazo de H_0. Sobre una base más teórica, puede demostrarse que

$$E(\text{CMR}) = \sigma^2 + \beta_1^2 \sum (x_i - \bar{x})^2$$

y

$$E(\text{CME}) = \sigma^2.$$

Si H_0 es cierta, entonces el valor esperado de *CMR* también es σ^2. Pero si $\beta_1 \neq 0$, E(CMR) es mayor que σ^2, ya que el término $\beta_1^2 \, \Sigma(x_i - \bar{x})^2$ es positivo. Por lo tanto, la estadística apropiada está dada por (13.23) con el extremo superior de la distribución F como región crítica; es decir, para un tamaño dado del error de tipo I α se rechaza la hipótesis nula de no regresión lineal cuando un valor de $F = CMR/CME$ se encuentra dentro de la región crítica superior de la distribución F con 1 y n -2 grados de libertad. La tabla de análisis de varianza (ANOVA) para el modelo lineal simple se encuentra en la tabla 13.5.

Para calcular las sumas de cuadrados que aparecen en la tabla 13.5, se tiene

$$\text{STC} = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n},$$

$$\text{SCE} = \sum (y_i - \hat{y}_i)^2 = \sum e_i^2,$$

donde $y_1, y_2, \ldots, y_n$ son las verificaciones de las observaciones, y $e_1, e_2, \ldots, e_n$ son los residuos correspondientes. Entonces $SCR = STC - SCE$, o puede calcularse

TABLA 13.5 Tabla ANOVA para el modelo lineal simple

Fuente de variación	*gl*	SC	CM	*Estadística F*
Regresión	1	$\sum (\hat{Y}_i - \bar{Y})^2$	$\sum (\hat{Y}_i - \bar{Y})^2/1$	$\dfrac{\sum (\hat{Y}_i - \bar{Y})^2/1}{\sum (Y_i - \hat{Y}_i)^2/(n-2)}$
Error	$n-2$	$\sum (Y_i - \hat{Y}_i)^2$	$\sum (Y_i - \hat{Y}_i)^2/(n-2)$	
Total	$n-1$	$\sum (Y_i - \bar{Y})^2$		

en forma directa. Dado que la recta estimada es

$$\hat{y}_i = \bar{y} + b_1(x_i - \bar{x})$$

o

$$\hat{y}_i - \bar{y} = b_1(x_i - \bar{x}),$$

al elevar al cuadrado ambos miembros y si se suman para todas las $i = 1, 2 \ldots n$ se obtiene

$$\text{SCR} = \sum (\hat{y}_i - \bar{y})^2 = b_1^2 \sum (x_i - \bar{x})^2. \tag{13.24}$$

Como ejemplo, recuérdese de nuevo el problema de los salarios inciales. Supóngase que se desea probar la hipótesis nula de que no existe una regresión lineal entre el salario inicial y *CP*, contra la alternativa de que ésta existe, con $\alpha = 0.01$. Mediante el uso de la tabla 13.2 se calculan las siguientes cantidades:

$$\text{STC} = 4970.12 - \frac{(270.8)^2}{15} = 81.2773,$$

$$\text{SCE} = 22.8671,$$

$$\text{SCR} = 81.2773 - 22.8671 = 58.4102.$$

Para $n = 15$ se proporciona la tabla ANOVA en la tabla 13.6. Dado que $f = 33.21 > f_{0.99,\,1,\,13} = 9.07$, se rechaza la hipótesis nula de no regresión lineal y se concluye que el salario inicial promedio está influenciado, en forma lineal, por la calificación promedio.

Como es de esperarse, existe una relación entre la estadística F anterior con 1 y $n - 2$ grados de libertad y la correspondiente estadística t de Student (véase la sección 13.3) para una hipótesis alternativa bilateral. Puede establecerse la relación mediante lo siguiente: dado que

$$\text{SCR} = B_1^2 \sum (x_i - \bar{x})^2$$

y

$$s^2(B_1) = \text{CME} \Big/ \sum (x_i - \bar{x})^2,$$

TABLA 13.6 Tabla ANOVA para los salarios iniciales

Fuente de variación	gl	SC	CM	*Valor F*
Regresión	1	58.4102	58.4102	33.21
Error	13	22.8671	1.759	
Total	14	81.2773	$f_{0.99,\,1,\,13} = 9.07$	

entonces

$$F = \frac{\text{CMR}}{\text{CME}} = \frac{B_1^2 \sum (x_i - \bar{x})^2/1}{s^2(B_1) \sum (x_i - \bar{x})^2} = [B_1/s(B_1)]^2.$$

De acuerdo con lo anterior, si una variable aleatoria tiene una distribución F con 1 y $n-2$ grados de libertad, entonces

$$F = T^2,$$

donde T es una variable aleatoria t de Student con $n-2$ grados de libertad. La relación entre los cuantiles es

$$f_{1-\alpha,\,1,\,n-2} = t^2_{1-\alpha/2,\,n-2}. \tag{13.25}$$

Hasta aquí se han examinado algunas maneras para probar la hipótesis nula de no regresión lineal entre x y Y. Ahora se presentará una cantidad numérica muy útil que es una medida relativa del grado de asociación lineal entre x y Y. Lo que se desea es tener una cantidad que mida la proporción de la variación total de las observaciones con respecto a su media la cual es atribuida a la recta estimada de regresión. Dado que *STC* representa la variación total con respecto a la media y *SCR* mide la porción de ésta, que es atribuible a un efecto lineal de x sobre Y, una medida apropiada es

$$r^2 = \frac{\text{SCR}}{\text{STC}} = \frac{\text{STC} - \text{SCE}}{\text{STC}} = 1 - \frac{\text{SCE}}{\text{STC}}. \tag{13.26}$$

r^2 recibe el nombre de *coeficiente de determinación*. Los valores que toma están siempre en el intervalo $0 \leq r^2 \leq 1$ ya que $0 \leq \text{SCE} \leq \text{STC}$. De manera ideal, se desea tener un $r^2 = 1$ ya que entonces $SCE = 0$, y toda la variación presente en las observaciones puede explicarse por la presencia lineal de x en la ecuación de regresión. De esta forma, entre más cercano se encuentre r^2 a uno, mayor es el grado de asociación lineal que existe entre x y Y. Como ilustración, el coeficiente de determinación para el ejemplo del salario inicial, es

$$r^2 = 1 - \frac{22.8671}{81.2773} = 0.7187.$$

Por lo tanto, la presencia lineal de *CP* en el modelo de regresión explica el 71.87% de la variación total en los salarios iniciales observados.

Ya que muchas veces se da una mala interpretación a r^2, debe hacerse un comentario sobre lo que r^2 no mide. r^2 no mide la validez del modelo de regresión propuesto, es decir, r^2 no puede verificar que la verdadera ecuación de regresión entre x y Y sea estrictamente lineal. Todo lo que puede medir es cuánto se explica de la variación total mediante la ecuación de regresión estimada. En realidad, el modelo verdadero de regresión entre x y Y puede contener términos no lineales en x, u otras variables de predicción, o ambos. Estas cuestiones serán examinadas en el capítulo 14.

A continuación se presenta una muestra de un listado de computadora para el análisis de regresión lineal de los datos de salarios. La lista de paquetes estadísticos disponible para computadora incluye a *SAS, SPSS, BMDP* y Minitab. El listado que se muestra en la figura 13.4 fue generado por Minitab. Nótese que incluye los coeficientes de la regresión estimados, sus desviaciones estándar, la prueba T para pendiente cero, la desviación estándar residual (o la desviación estándar de Y con respecto a la recta de regresión); el valor de r^2 las sumas de los cuadrados, los cuadrados medios para el análisis de varianza y los residuos estandarizados definidos en el capítulo 12.

LA ECUACIÓN DE REGRESIÓN ES
Y = − 6.63 + 8.12 X1

	COLUMNA	COEFICIENTE	DEV. EST. DEL COEF.	COCIENTE-T = COEF/D.E.
	—	− 6.627	4.298	−1.54
X1	C2 ·	8.118	1.409	5.76

LA DEV. EST. DE Y CON RESPECTO A LA RECTA DE REGRESIÓN ES
S = 1.327
CON (15− 2) = 13 GRADOS DE LIBERTAD

R-CUADRADO = 71.8%

ANÁLISIS DE VARIANZA

DEBIDA A	DF	SC	CM = SC/GL
REGRESIÓN	1	58.393	58.393
RESIDUO	13	22.880	1.760
TOTAL	14	81.274	

RENGLON	X1 C2	Y C1	VALOR PRED. Y	DEV. EST. PRED. Y	RESIDUO	RES. EST.
1	2.95	18.500	17.323	0.365	1.177	0.92
2	3.20	20.000	19.352	0.410	0.648	0.51
3	3.40	21.100	20.976	0.612	0.124	0.11
4	3.60	22.400	22.600	0.860	−0.200	−0.20
5	3.20	21.200	19.352	0.410	1.848	1.46
6	2.85	15.000	16.511	0.435	−1.511	−1.21
7	3.10	18.000	18.540	0.353	−0.540	−0.42
8	2.85	18.800	16.511	0.435	2.289	1.83
9	3.05	15.700	18.134	0.343	−2.434	−1.90
10	2.70	14.400	15.293	0.589	−0.893	−0.75
11	2.75	15.500	15.699	0.533	−0.199	−0.16
12	3.10	17.200	18.540	0.353	−1.340	−1.05
13	3.15	19.000	18.946	0.376	0.054	0.04
14	2.95	17.200	17.323	0.365	−0.123	−0.10
15	2.75	16.800	15.699	0.533	1.101	0.91

FIGURA 13.4 Listado de computadora para el análisis de regresión lineal (datos de los salarios iniciales)

13.8 Correlación lineal

En la sección 6.4 se definió el coeficiente de correlación ρ dado por (6.14), como una medida de la asociación lineal que existe entre las variables aleatorias X y Y. En esta sección se examinará el coeficiente de correlación de la muestra en el contexto del análisis de regresión.

Durante toda la presentación del análisis de regresión se ha asumido la disponibilidad de una muestra aleatoria de la variable respuesta $Y_1, Y_2, \ldots, Y_n$, correspondientes a n valores fijos $x_1, x_2, \ldots, x_n$ de una variable de predicción. Para definir el coeficiente de correlación de la muestra, se supondrá que tanto X como Y son variables aleatorias. Sea la distribución conjunta de X y Y la normal bivariada (véase la sección 6.8), y sean $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ una muestra aleatoria de tamaño n de esta distribución. Entonces puede demostrarse que el estimador de máxima verosimilitud de ρ (denominado ***coeficiente de correlación de la muestra***), está dado por

$$r^*(X, Y) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\left[\sum_{i=1}^{n} (X_i - \bar{X})^2\right]^{1/2} \left[\sum_{i=1}^{n} (Y_i - \bar{Y})^2\right]^{1/2}}. \tag{13.27}$$

Después de efectuar algunos cálculos algebraicos, puede obtenerse una expresión equivalente de la forma

$$r(X, Y) = \frac{\sum X_i Y_i - \dfrac{\left(\sum X_i\right)\left(\sum Y_i\right)}{n}}{\left[\sum X_i^2 - \dfrac{\left(\sum X_i\right)^2}{n}\right]^{1/2} \left[\sum Y_i^2 - \dfrac{\left(\sum Y_i\right)^2}{n}\right]^{1/2}}. \tag{13.28}$$

Al igual que el parámetro ρ, r se encuentra en el intervalo $-1 \leq r \leq 1$ y mide la relación lineal entre X y Y, si X se emplea para predecir Y o viceversa. Con base en una muestra aleatoria, un valor de $r = -1$ indica una relación lineal negativa perfecta entre X y Y, mientras que un valor de $r = 1$ señalará una asociación lineal positiva perfecta de X y Y. Si $r = 0$, entonces no existe ninguna relación lineal entre X y Y. En la figura 13.5 se muestran algunas gráficas de dispersión comunes para algunos valores de r.

A causa de varias interpretaciones injustificables que ha sufrido r, es imperioso que el lector comprenda que r por sí mismo no puede ni probar ni desmentir una relación causal entre X y Y, aun si $r = \pm 1$. Como ya se indicó al principio de este capítulo, la manifestación de una relación causa-efecto es posible sólo a través de la comprensión de la relación natural que existe entre X y Y, y ésta no debe manifestarse sólo por la existencia de una fuerte correlación entre X y Y.

* Se seguirá la norma de utilizar una r minúscula.

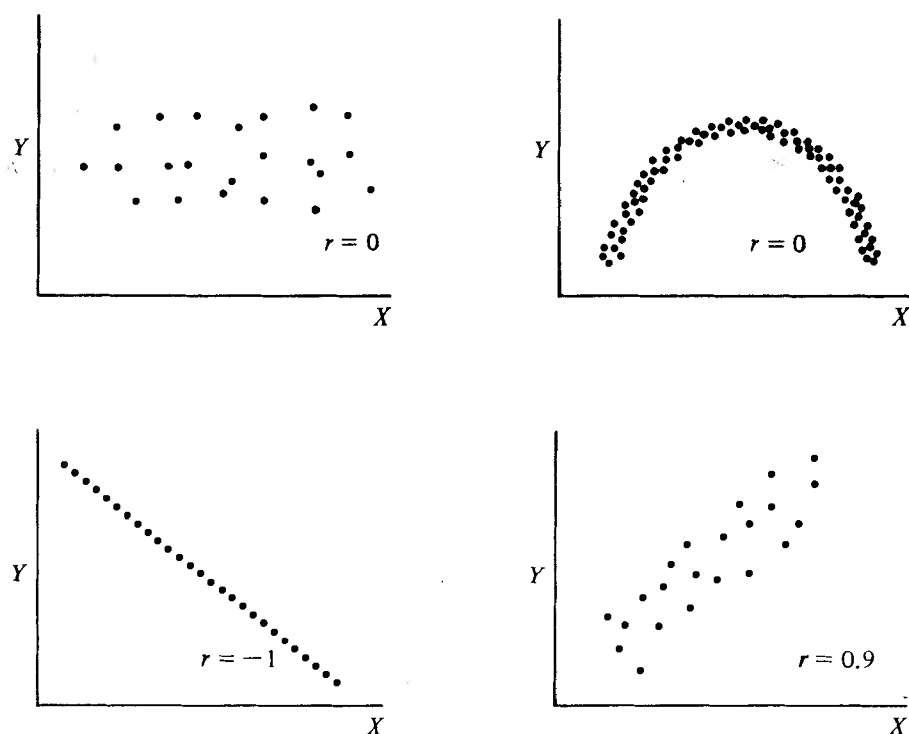


FIGURA 13.5 Gráficas de dispersión comunes para algunos valores de r

Mientras que en el análisis de regresión se supone que los valores de x son fijos, el coeficiente de correlación de la muestra definido por (13.27) o (13.28) es todavía un estimador de ρ. Dado que r mide el grado de asociación lineal entre x y Y, y ya que B_1 es el correspondiente estimador de mínimos cuadrados de la pendiente para el modelo lineal propuesto entre x y Y, entonces debe existir una relación entre r y B_1. Mediante el empleo de la segunda ecuación de (13.6) y (13.27), puede demostrarse que el estimador de mínimos cuadrados de la pendiente y el correspondiente valor del coeficiente de correlación de la muestra se encuentran relacionados por

$$b_1 = \left[\frac{\sum (y_i - \bar{y})^2}{\sum (x_i - \bar{x})^2}\right]^{1/2} r. \qquad (13.29)$$

Nótese que si $r = 0$, $b_1 = 0$ y viceversa. Además, el signo de b_1 siempre es igual al de r. Finalmente, el cuadrado del coeficiente de correlación de la muestra es el coeficiente de determinación, es decir, si r^2 y b_1 son conocidos, entonces se sabe el valor de r y su signo; por lo tanto, se sigue que r no sólo es una medida del grado de aso-

ciación lineal entre dos variables, sino que puede emplearse una función de r como una medida de la bondad del ajuste para una ecuación estimada de regresión.

13.9 Series de tiempo y autocorrelación

En las secciones anteriores se han examinado los análisis de regresión y de correlación con base en una muestra aleatoria de la variable respuesta Y. En muchas situaciones, por ejemplo en economía y finanzas, la variable respuesta se mide en forma periódica con respecto al tiempo. Por ejemplo, puede escogerse examinar la tasa de desempleo para los pasados 24 meses, o puede observarse el volumen de ventas trimestral de alguna compañía y compararlo con el correspondiente volumen de ventas de toda la industria durante los pasados 12 trimestres. Dado que para ambos ejemplos las observaciones se registran de manera secuencial con el paso del tiempo, forman lo que se conoce como una *serie de tiempo*.

Aunque los metodos de regresión pueden ser útiles al analizar datos de series de tiempo, las observaciones de Y en una serie de tiempo no pueden considerarse como representativas de una muestra aleatoria. De hecho, pueden encontrarse correlacionadas entre sí. Por ejemplo, es probable que el cambio en la tasa de desempleo para este mes se encuentre relacionada con la que se observará para el siguiente mes. De esta forma, algunas de las suposiciones que son necesarias para el desarrollo de procedimientos inferenciales posiblemente no se verifiquen para los datos de una serie de tiempo.

En este contexto se desea considerar un procedimiento inferencial útil, conocido como estadística de Durbin-Watson, para determinar si los errores en un modelo* lineal sencillo se encuentran correlacionados en el tiempo. Los errores del mismo modelo de regresión que se encuentran correlacionados como funciones del tiempo reciben el nombre de *correlacionados serialmente* o *autocorrelacionados*. Antes de analizar el procedimiento de Durbin-Watson, se mencionarán en forma breve los componentes usuales de datos de una serie de tiempo.

13.9.1 Componentes de una serie de tiempo

Las fluctuaciones de la variable respuesta en una serie de tiempo de tipo económico se asignan, por lo general, a cuatro causas diferentes (componentes): la variación en la tendencia T, la variación por temporada S, la variación cíclica C y la variación aleatoria R. la *variación en la tendencia* es el movimiento a largo plazo en Y. Por ejemplo, la producción de automóviles en Estados Unidos ha mostrado una tendencia hacia el crecimiento durante los últimos 50 años, pero lo anterior no necesariamente implica que la producción aumentó todos los años durante este periodo. De esta forma, la tendencia refleja el movimiento general de Y a lo largo de un periodo. La *variación por temporada* representa el movimiento de Y que ocurre durante periodos específicos a lo largo de un año. Por ejemplo, el volumen de ventas al menudeo tiende a ser mayor en el último trimestre del año que durante el primero. La *va-*

* También puede emplearse este procedimiento para el modelo lineal general, el cual se estudiará en el capítulo 14.

riación cíclica muestra el movimiento de Y que se repite durante periodos que, en general, son mayores de un año. Los movimientos cíclicos se encuentran muchas veces relacionados con las condiciones económicas prevalecientes. Por ejemplo, la construcción de casas en Estados Unidos disminuyó durante el periodo de recesión de 1974-1975; aumentó durante el de recuperación de 1976-1979 y volvió a disminuir en la recesión de 1981-1982. La *variación aleatoria* en una serie de tiempo es la fluctuación de Y que no es posible asignar a una causa identificable. Por lo tanto, la fluctuación total de Y con respecto al tiempo se asigna a una variación sistemática (tendencia, temporada y ciclo) y a una variación aleatoria.

Al suponer cómo se encuentran relacionadas estas componentes, puede formularse un modelo de una serie de tiempo que ayudará a separar estas componentes y formular predicciones con respecto a Y. Los modelos de las series de tiempo usualmente son aditivos de la forma

$$Y = T + S + C + R$$

o multiplicativos de la forma

$$Y = T \times S \times C \times R.$$

Para un modelo aditivo se supone que los cuatro componentes son independientes entre sí, mientras que para el multiplicativo se encuentran relacionados entre sí. Para tratamientos completos del análisis de las series de tiempo se sugieren las referencias [1] y [4].

13.9.2 La estadística de Durbin-Watson

En esta sección el interés radicará, en forma exclusiva, en la detección de errores autocorrelacionados y en un análisis con respecto a medidas correctivas. Una de las razones de la existencia de la autocorrelación es que podrían no haberse tomado en cuenta en el modelo variables importantes de predicción. Por ejemplo, se mencionó que, en general, la producción de automóviles tuvo un incremento durante un periodo de 50 años. Si se supone algún modelo de regresión con el tiempo como la única variable de predicción, no es de dudar que se encontrarán correlaciones entre los errores. Pero durante el mismo periodo aumentó la población así como el nivel económico de los habitantes de Estados Unidos. Cuando variables de predicción como éstas están positivamente correlacionadas con la producción de automóviles, pero no se toman en cuenta en el modelo de regresión, entonces los errores tenderán a estar positivamente autocorrelacionados, ya que también reflejan los efectos de las variables de predicción faltantes. Este tipo de autocorrelación sólo es aparente y puede eliminarse mediante la inclusión de las variables omitidas en el modelo de regresión.

En las series de tiempo económicas, la autocorrelación también puede presentarse debido a que los residuos sucesivos tienden a estar positivamente correlacionados, es decir, los grandes residuos negativos siguen a grandes residuos negativos y los grandes residuos positivos siguen a grandes residuos positivos. Este tipo de autocorrelación es, en general, la clase que necesita algún ajuste. El interés recaerá en este tipo y se estudiarán las medidas correctivas tales como la transformación de los datos.

Recuérdese que por la suposición 3 de la sección 13.2, la covarianza entre los errores aleatorios ε_i y ε_j es cero para todo $i \neq j$. A pesar de que esta suposición no es necesaria para obtener los estimadores de mínimos cuadrados, su violación afecta las propiedades inferenciales de estos estimadores. Cuando se encuentra presente la autocorrelación, el análisis de regresión es afectado en tres formas.

1. Los estimadores *MC*, aunque son no sesgados ya no tienen varianza mínima.
2. Los estimados $s^2(B_i)$ pueden subestimar, en forma seria, las varianzas de los estimadores *MC* de B_i.
3. Los intervalos de confianza y las pruebas de hipótesis que incluyen, ya sea la distribución t de Student o la distribución F, no son teóricamente válidas.

Por ejemplo, supóngase que los datos que figuran más adelante representadas las ventas Y de alguna compañia (en millones de dólares) y las ventas x (también en millones de dólares) para toda la industria en los pasados 16 trimestres, dond lo datos ya se han ajustado de acuerdo con la inflación.

t	1	2	3	4	5	6	7	8
x_t	270.36	258.38	254.96	259.70	265.40	274.98	281.86	285.78
Y_t	44.84	42.97	41.98	42.75	43.95	45.65	46.87	47.35

t	9	10	11	12	13	14	15	16
x_t	290.58	290.18	296.72	292.32	301.72	305.42	314.96	321.10
Y_t	48.13	47.95	49.10	48.52	50.22	51.15	52.78	53.91

Una gráfica de Y contra x revela una tendencia lineal, lo que a su vez sugiere que las acciones de la compañia se mantienen en el mercado. Supóngase un modelo lineal simple como el dado por (13.1). El listado de computadora producido por Minitab, se muestra en la figura 13.6

Nótese que parece que el modelo ajusta los datos en forma excelente, ya que $r^2 = 0.997$, y se rechaza la hipótesis nula de pendiente igual a cero para casi cualquier nivel α. Las desviaciones estádar estimadas para B_0 y B_1 son pequeñas, y en forma especial para el estimador de la pendiente. Pero al graficar los residuos estandarizados contra el tiempo, como se muestra en la figura 13.7, se nota que los residuos del mismo signo aparecen agrupados. Por ejemplo, los residuos 5-7 son positivos, 8-13 son negativos y 14-16 son positivos. Este tipo de patrón es característico cuando se tienen errores autocorrelacionados.

La estadística de Durbin-Watson constituye un enfoque más formal que al graficar los residuos para detectar los errores autocorrelacionados; se basa en la suposición de que los errores ε_t en el modelo de regresión

$$Y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \tag{13.30}$$

forman una serie autorregresiva de primer orden dada por

$$\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t \qquad t \geq 2, \tag{13.31}$$

LA ECUACION DE REGRESION ES
Y = − 2.97 + 0.177 X1

	COLUMNA	COEFICIENTE	DEV. EST. DEL COEF.	COCIENTE-T = COEF/D.E.
	–	−2.9716	0.7023	−4.23
X1	C2	0.176510	0.002456	71.86

LA DEV. EST. DE Y CON RESPECTO A LA RECTA DE REGRESION ES
S = 0.1919
CON (16 - 2) = 14 GRADOS DE LIBERTAD

R-CUADRADO = 99.7%

ANALISIS DE VARIANZA

DEBIDO A	DF	SC	CM = SC/GL
REGRESION	1	190.2330	190.2330
RESIDUO	14	0.5157	0.0368
TOTAL	15	190.7487	

RENGLON	X1 C2	Y C1	VALOR PRED. Y	DEV. EST. PRED. Y	RESIDUO	RES. EST.
1	270	44.8400	44.7497	0.0604	0.0903	0.50
2	258	42.9699	42.6350	0.0816	0.3349	1.93
3	255	41.9800	42.0314	0.0886	−0.0515	−0.30
4	260	42.7499	42.8680	0.0790	−0.1181	−0.68
5	265	43.9499	43.8742	0.0684	0.0758	0.42
6	275	45.6500	45.5651	0.0542	0.0848	0.46
7	282	46.8699	46.7795	0.0487	0.0904	0.49
8	286	47.3499	47.4714	0.0480	−0.1215	−0.65
9	291	48.1299	48.3187	0.0497	−0.1887	−1.02
10	290	47.9499	48.2481	0.0495	−0.2981	−1.61
11	297	49.0999	49.4025	0.0556	−0.3025	−1.65
12	292	48.5200	48.6258	0.0510	−0.1059	−0.57
13	302	50.2199	50.2850	0.0627	−0.0651	−0.36
14	305	51.1500	50.9381	0.0689	0.2119	1.18
15	315	52.7800	52.6220	0.0873	0.1579	0.92
16	321	53.9100	53.7058	0.1002	0.2042	1.25

FIGURA 13.6 Análisis de regresión lineal (datos del mercado de acciones)

donde $|\rho| < 1$ es la pendiente de la recta que pasa por el origen y η_t es el error aleatorio puro que no se encuentra correlacionado con cualquier otra componente. El término η_t se denomina de manera común como *ruido blanco*. Debe notarse que (13.31) es un modelo autorregresivo, ya que la variable de predicción ε_{t-1} es un término retardado en el tiempo de la variable respuesta ε_t. A pesar de que la estructura de correlación entre los errores puede ser más compleja que la implicada por (13.31), un modelo autorregresivo de primer orden es una aproximación razonable, debido a que muchas veces la autocorrelación entre ε_t y ε_{t+p} disminuye de manera rápida conforme la distancia entre los puntos en el tiempo t y $t + p$ aumenta.

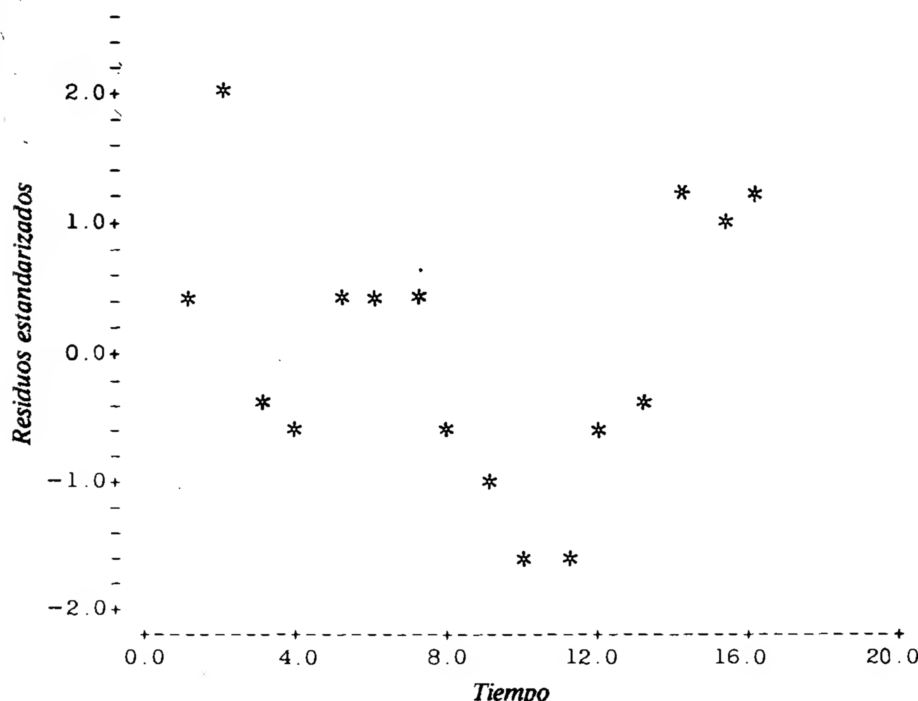


FIGURA 13.7 Residuos estandarizados contra tiempo para el ejemplo de las ventas

Para el modelo dado por (13.31), se desea emplear la estadística de Durbin-Watson para probar la hipótesis nula

$$H_0\colon \rho = 0$$

contra la alternativa

$$H_1\colon \rho > 0.$$

Nótese que H_1 es una hipótesis alternativa unilateral superior, ya que las series de tiempo económicas exhiben muchas veces una autocorrelación positiva. La estadística de Durbin-Watson se basa en los residuos que resultan después de obtener la ecuación de regresión estimada para (13.30). Se calcula un valor de esta estadística a partir de la expresión

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}, \qquad (13.32)$$

donde el residuo es $e_t = y_t - \hat{y}_t$.

Si los errores se encuentran positivamente autocorrelacionados, es probable que los errores adyacentes tengan la misma magnitud. De esta forma, pequeñas diferencias entre los residuos adyacentes sugieren que ρ es mayor que cero; pero cuando las diferencias son pequeñas, el numerador de (13.32) también lo es. De acuerdo con lo anterior, se rechaza la hipótesis nula de autocorrelación cero siempre que d tiene un valor relativamente pequeño.

Durbin y Watson tabularon los límites inferior y superior d_L y d_U, respectivamente, para probar H_0. En la tabla K del apéndice se proporcionan los límites d_L y d_U para $\alpha = 0.05$ y 0.01 como funciones del tamaño n de la muestra y el número k de variables de predicción en el modelo de regresión. Dados los límites d_L y d_U, la decisión para H_0 se toma de la siguiente forma:

a Si $d < d_L$, rechazar H_0,

b Si $d > d_U$, no puede rechazarse H_0,

c Si $d_L < d < d_U$, la prueba no es concluyente.

Debe señalarse que la prueba para autocorrelación negativa (H_1: $\rho < 0$) también es posible con la estadística de Durbin-Watson. En este caso, el valor de la estadística es $4 - d$, donde d se calcula de acuerdo con (13.32). El procedimiento de decisión es igual al ya dado, comparando $4 - d$ con d_L o d_U. En cualquier caso, si la prueba es no concluyente, la alternativa que se sugiere es tomar más observaciones.

Para el ejemplo se calcula d primero, con lo que se obtienen las diferencias $e_t - e_{t-1}$, mediante el uso de la columna de residuos dada en el listado de computadora. Estas diferencias son las siguientes:

t	2	3	4	5	6
$e_t - e_{t-1}$	0.2446	−0.3864	−0.0666	0.1939	0.0090

t	7	8	9	10	11
$e_t - e_{t-1}$	0.0056	−0.2119	−0.0672	−0.1094	−0.0044

t	12	13	14	15	16
$e_t - e_{t-1}$	0.1966	0.0408	0.2770	−0.0540	0.0463

Mediante el empleo de (13.32) se obtiene

$$d = 0.434789/0.5157 = 0.843.$$

Por ejemplo, $\alpha = 0.05$; entonces para el modelo lineal simple (13.30) y $n = 16$, los límites son $d_L = 1.10$ y $d_U = 1.37$. Dado que $d < d_L$, se rechaza la hipótesis nula y se concluye que existe una razón para creer que los errores en (13.30) se encuentran autocorrelacionados.

13.9.3 Eliminación de la autocorrelación mediante la transformación de datos

Cuando se rechaza la hipótesis nula de autocorrelación cero, debe ajustarse la ecuación estimada de regresión para compensar la presencia de errores autocorrelacionados. A continuación se mostrará un enfoque debido a Cochrane y Orcutt.* Se basa en un método iterativo el cual incluye la transformación de las variables respuesta y predicción en el modelo original de regresión.

Para el modelo dado por (13.30), considérese la transformación

$$Y_t' = Y_t - \rho Y_{t-1}. \tag{13.33}$$

Al sustituir en (13.33) Y_t y Y_{t-1}, de acuerdo con (13.30), se tiene que

$$\begin{aligned} Y_t' &= (\beta_0 + \beta_1 x_t + \varepsilon_t) - \rho(\beta_0 + \beta_1 x_{t-1} + \varepsilon_{t-1}) \\ &= \beta_0(1 - \rho) + \beta_1(x_t - \rho x_{t-1}) + (\varepsilon_t - \rho\varepsilon_{t-1}). \end{aligned}$$

Pero de (13.31)

$$\varepsilon_t - \rho\varepsilon_{t-1} = \eta_t$$

donde η_t son errores aleatorios no correlacionados. Entonces

$$Y_t' = \beta_0(1 - \rho) + \beta_1(x_t - \rho x_{t-1}) + \eta_t,$$

o

$$Y_t' = \beta_0' + \beta_1' x_t' + \eta_t, \tag{13.34}$$

donde $\beta_0' = \beta_0(1 - \rho)$, $\beta_1' = \beta_1$, y $x_t' = x_t - \rho x_{t-1}$. De acuerdo con lo anterior, los errores en el modelo lineal simple transformado (13.34) no están correlacionados entre sí, y de esta forma este modelo satisface las suposiciones estándar.

Nótese que las observaciones transformadas $Y_t' = Y_t - \rho Y_{t-1}$ y $x_t' = x_t - \rho x_{t-1}$ son funciones de la autocorrelación desconocida ρ, así que antes de ajustar el modelo transformado debe obtenerse un estimador de ρ. Lo anterior puede hacerse mediante el empleo de los residuos obtenidos de la ecuación de regresión estimada originalmente para calcular un estimador *MC* de la pendiente ρ en el modelo autorregresivo de primer orden dado por (13.31). Ya que este modelo tiene una intersección igual a cero, el estimador *MC*, r de la pendiente ρ basado en el análisis de la sección 13.3, es

$$r = \frac{\sum_{t=2}^{n} e_{t-1}e_t}{\sum_{t=1}^{n} e_t^2}, \tag{13.35}$$

*D. Cochrane y G. H. Orcutt, *Application of least squares regression to relationships containing autocorrelated error terms*, J. Amer. Statistical Assoc. **44** (1949), 32-61.

y los valores transformados son

$$y'_t = y_t - ry_{t-1},$$
$$x'_t = x_t - rx_{t-1}. \quad (13.36)$$

Dados los valores transformados para las variables de respuesta y predicción, el procedimiento iterativo consiste en determinar la ecuación de regresión estimada para el modelo transformado y entonces volver a calcular la estadística de Durbin-Watson. Si no es posible rechazar la hipótesis nula de autocorrelación cero, el procedimiento llega a su fin. De otra forma, se repite hasta que H_0 no pueda rechazarse. Si se requiere más de una iteración, entonces se sugiere buscar otros procedimientos alternativos.

Como ejemplo, el estimado *MC* de ρ para el ejemplo de las ventas es

$$r = 0.2734/0.5157 = 0.53,$$

y los valores transformados son los siguientes:

t	2	3	4	5	6	7	8	9
x'_t	115.09	118.02	124.57	127.76	134.32	136.12	136.39	139.12
Y'_t	19.20	19.21	20.50	21.29	22.36	22.68	22.51	23.03

t	10	11	12	13	14	15	16
x'_t	136.17	142.92	135.06	146.79	145.51	153.09	154.17
Y'_t	22.44	23.69	22.50	24.50	24.53	25.67	25.94

El listado de computadoras que se obtiene mediante el empleo de Minitab* para el modelo transformado se muestra en la figura 13.8. Nótese que el listado también incluye el valor de $d = 1.61$ para la estadística de Durbin-Watson; Minitab proporciona este valor como parte del listado. Para $n = 15$ y $\alpha = 0.05$, se obtienen los límites $d_L = 1.08$ y $d_U = 1.36$ al consultar la tabla *K*. Dado que $d > d_U$, no es posible rechazar la hipótesis nula de autocorrelación cero.

Ahora, es necesario escribir la ecuación de regresión estimada en términos de las variables originales y ajustar las desviaciones estándar estimadas de B_0 y B_1 para reflejar la eliminación de los errores autocorrelacionados. Dado que $\beta'_0 = \beta_0(1 - \rho)$ y $\beta'_1 = \beta_1$, los estimadores *MC* de β_0 y β_1 son

$$b_0 = \frac{b'_0}{(1 - r)} = \frac{-1.5178}{(1 - 0.53)} = -3.2294,$$

y

$$b_1 = b'_1 = 0.1774.$$

* Se ha omitido una porción del listado que incluye los valores de las variables de respuesta y predicción, residuos, etc.

LA ECUACION DE REGRESION ES
Y = −1.52 + 0.177 X1

	COLUMNA	COEFICIENTE	DEV. EST. DEL COEF.	COCIENTE-T = COEF/D.E.
	–	−1.5178	0.5176	−2.93
X1	C2	0.177407	0.003784	46.88

LA DEV. EST. DE T CON RESPECTO A LA RECTA DE REGRESION ES
S = 0.1627
CON (15 − 2) = 13 GRADOS DE LIBERTAD

R-CUADRADO = 99.4%

ANALISIS DE VARIANZA

DEBIDA A	DF	SC	CM = SC/GL
REGRESION	1	58.16086	58.16086
RESIDUO	13	0.34401	0.02646
TOTAL	14	58.50485	

ESTADISTICA DE DURBIN-WATSON = 1.61

FIGURA 13.8 Análisis de regresión lineal después de la transformación de los datos por autocorrelación

Para los estimadores B_0' y B_1' del listado de la figura 13.8, se nota que sus desviaciones estándar estimadas son $s(B_0') = 0.5176$ y $s(B_1') = 0.003784$. Por lo tanto, para las desviaciones estándar estimadas de B_0 y B_1, se tiene

$$s(B_0) = s\left[\frac{B_0'}{(1 - r)}\right] = s(B_0')/(1 - r) = 1.1013,$$

$$s(B_1) = s(B_1') = 0.003784.$$

En la tabla 13.7 se encuentra un resumen de la información pertinente para las ecuaciones de regresión estimadas original y final para los datos de ventas. Nótese que a pesar de que el cambio en los valores estimados de los coeficientes es pequeño, existe un considerable aumento en las desviaciones estándar estimadas de B_0 y B_1, y en

TABLA 13.7 Resumen de la información para los datos de ventas

Ecuación original estimada	*Ecuación final estimada*
$\hat{y}_t = -2.9716 + 0.1765x_t$	$\hat{y}_t = -3.2294 + 0.1774x_t$
$s(B_0) = 0.7023, \quad s(B_1) = 0.002456$	$s(B_0) = 1.1013, \quad s(B_1) = 0.003784$
CME = 0.0368	CME = 0.0265
$r^2 = 0.997$	$r^2 = 0.994$

forma especial para B_1. Pero la varianza residual (*CME*) ha disminuido. En este ejemplo, la autocorrelación aparente no fue lo suficientemente fuerte como para causar diferencias sustanciales en la inferencia. Cuando ocurre lo contrario, es probable que se noten diferencias muy drásticas.

13.10 Enfoque matricial para el modelo lineal simple

El uso del álgebra de matrices proporciona un medio conveniente para el análisis de regresión de modelos lineales, en forma especial de aquellos que contienen más de una variable de predicción. Se ilustrará el uso del álgebra de matrices mediante el examen del modelo lineal simple. Para una breve revisión de los fundamentos del álgebra de matrices, se invita al lector a que consulte el apéndice que se encuentra al final de este capítulo.

Para los n pares $(x_1, Y_1), (x_2, Y_2), \ldots, (x_n, Y_n)$, el siguiente modelo lineal simple

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad i = 1, 2, \ldots, n.$$

En otras palabras,

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 x_1 + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 x_2 + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 x_n + \varepsilon_n \end{aligned}$$

son n ecuaciones lineales para las que $Y_1, Y_2, \ldots, Y_n$ son las observaciones de la respuesta para los correspondientes valores fijos $x_1, x_2, \ldots, x_n$ de la variable de predicción, $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ son los errores aleatorios no observables y β_0 y β_1 son los parámetros por estimarse. Si se definen las matrices

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

entonces

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 x_1 + \varepsilon_1 \\ \beta_0 + \beta_1 x_2 + \varepsilon_2 \\ \vdots \\ \beta_0 + \beta_1 x_n + \varepsilon_n \end{bmatrix}.$$

Como resultado se tiene que el modelo lineal simple puede expresarse en la notación de matrices

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{13.37}$$

Si se supone el caso de la teoría normal, entonces $\boldsymbol{\varepsilon}$ es un vector de variables aleatorias normales, tal que

$$E(\boldsymbol{\varepsilon}) = \mathbf{0},$$

$$Var(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$$

donde σ^2 es la varianza del error, común a todos ellos, e $\mathbf{I}$ es la matriz de identidad correspondiente.

Ahora, considérese la estimación de mínimos cuadrados de β_0 y β_1. Recuérdese que las ecuaciones normales están dadas por (13.4). Dado que

$$\mathbf{X'X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix} \tag{13.38}$$

y

$$\mathbf{X'Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum x_i Y_i \end{bmatrix}, \tag{13.39}$$

entonces

$$(\mathbf{X'X})\mathbf{B} = \begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix} \begin{bmatrix} \mathrm{B}_0 \\ \mathrm{B}_1 \end{bmatrix} = \begin{bmatrix} nB_0 & + B_1\Sigma x_i \\ B_0\Sigma x_i & + B_1\Sigma x_i^2 \end{bmatrix} = \begin{bmatrix} \Sigma Y_i \\ \Sigma x_i Y_i \end{bmatrix} = \mathbf{X'Y}.$$

Por lo tanto, las ecuaciones normales en forma matricial son

$$(\mathbf{X'X})\mathbf{B} = \mathbf{X'Y}, \tag{13.40}$$

donde

$$\mathbf{B} = \begin{bmatrix} B_0 \\ B_1 \end{bmatrix}$$

es el vector que contiene los estimadores de mínimos cuadrados B_0 y B_1.

Si se supone que la matriz cuadrada $\mathbf{X'X}$ tiene inversa, entonces en (13.40)

$$(\mathbf{X'X})^{-1}(\mathbf{X'X})\mathbf{B} = (\mathbf{X'X})^{-1}\mathbf{X'Y},$$

o

$$\mathbf{IB} = (\mathbf{X'X})^{-1}\mathbf{X'Y},$$

y

$$\mathbf{B} = (\mathbf{X'X})^{-1}\mathbf{X'Y} \tag{13.41}$$

es la expresión matricial para obtener los estimadores de mínimos cuadrados B_0 y B_1.

Al emplear los datos correspondientes al ejemplo de salarios inciales, se ilustrará que la expresión dada por (13.41) proporciona los mismos estimadores de mínimos

cuadrados para β_0 y β_1 obtenidos con anterioridad. El vector **Y** de salarios iniciales y la matriz **X** correspondiente a las calificaciones promedio son

$$\mathbf{Y} = \begin{bmatrix} 18.5 \\ 20.0 \\ 21.1 \\ \vdots \\ 16.8 \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} 1 & 2.95 \\ 1 & 3.20 \\ 1 & 3.40 \\ \vdots & \vdots \\ 1 & 2.75 \end{bmatrix}.$$

El lector debe notar que los números uno que se encuentran en la primera columna de **X** representan la intersección β_0 definida de acuerdo con el modelo lineal simple propuesto. Al seguir con el cálculo se tiene

$$(\mathbf{X}'\mathbf{X}) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 2.95 & 3.20 & \cdots & 2.75 \end{bmatrix} \begin{bmatrix} 1 & 2.95 \\ 1 & 3.20 \\ \vdots & \vdots \\ 1 & 2.75 \end{bmatrix} = \begin{bmatrix} 15 & 45.6 \\ 45.6 & 139.51 \end{bmatrix}$$

y

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 2.95 & 3.20 & \cdots & 2.75 \end{bmatrix} \begin{bmatrix} 18.5 \\ 20.0 \\ \vdots \\ 16.8 \end{bmatrix} = \begin{bmatrix} 270.8 \\ 830.425 \end{bmatrix}$$

La inversa de la matriz de 2×2 es igual a

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{13.29} \begin{bmatrix} 139.51 & -45.6 \\ -45.6 & 15 \end{bmatrix}$$

Para evitar la posibilidad de graves errores por redondeo, lo mejor es no dividir cada elemento de $(\mathbf{X}'\mathbf{X})^{-1}$ por el valor 13.29 hasta que se efectúe el producto $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Entonces, de (13.41) los estimadores de mínimos cuadrados son

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \frac{1}{13.29} \begin{bmatrix} 139.51 & -45.6 \\ -45.6 & 15 \end{bmatrix} \begin{bmatrix} 270.8 \\ 830.425 \end{bmatrix}$$

$$= \frac{1}{13.29} \begin{bmatrix} -88.072 \\ 107.895 \end{bmatrix} = \begin{bmatrix} -6.6269 \\ 8.1185 \end{bmatrix},$$

o $b_0 = -6.6269$ y $b_1 = 8.1185$. Al redondear a dos dígitos significativos, estos valores son iguales a los ya obtenidos con anterioridad.

Referencias

1. G. E. P. Box and G. M. Jenkins, *Time series analysis: Forecasting and control*, 2nd ed., Holden–Day, San Francisco, 1977.
2. S. Chatterjee and B. Price, *Regression analysis by example*, Wiley, New York, 1977.
3. N. R. Draper and H. Smith, *Applied regression analysis*, 2nd ed., Wiley, New York, 1981.
4. C. R. Nelson, *Applied time series analysis for managerial forecasting*, Holden–Day, San Francisco, 1977.
5. J. Neter and W. Wasserman, *Applied linear statistical models*, Richard D. Irwin, Homewood, Ill., 1974.

Ejercicios

13.1. Formúlese un comentario con respecto a la causalidad para las siguientes situaciones:

a) Durante los pasados 12, de 15 años, el mercado de valores creció cuando el promedio global de la liga mayor de beisbol disminuyó y viceversa.
b) Desde el primer Super Tazón en 1967, el mercado de valores aumentó durante todos los años en los que el equipo que ganaba el Super Tazón provenía de la vieja Liga Nacional de futbol, y disminuyó durante los años en lo que el campeón era un equipo de la vieja Liga Americana de futbol.

13.2. De los siguientes modelos, ¿cuáles son lineales?

a. $Y = \beta \operatorname{sen}(x) + \varepsilon$
b. $Y = \beta_1 \operatorname{sen}(\beta_2 x) + \varepsilon$
c. $Y = \beta_0 + \beta_1 x_1^2 x_2 + \beta_2 x_2^3 + \varepsilon$
d. $Y = \beta_0 + \beta_1^2 x + \varepsilon$

13.3. Dado el modelo lineal $Y_i = \beta x_i + \varepsilon_i$, $i = 1, 2, \ldots, n$, supóngase que $E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma^2$ para toda i y $Cov(\varepsilon_i, \varepsilon_j) = 0$ para toda $i \neq j$.

a) Obténgase el estimador **B** de mínimos cuadrados para β.
b) Determínese si **B** es un estimador no sesgado de β, y demuéstrese que $Var(B) = \sigma^2 / \Sigma x_i^2$.

13.4. Una compañía local de energía seleccionó una residencia típica para desarrollar un modelo empírico para el consumo de energía (en kilowatts por día) como una función de la temperatura promedio diaria durante los meses de invierno. Se obtuvo la siguiente información durante un periodo de 15 días.

Temperatura (°C)	0	8	7.5	13.5	14	8.5	4.5	−11
Consumo de energía	70	57	60	63	57	66	67	107
Temperatura (°C)	−7.5	−8.5	1.5	0.5	2	−6	−4	
Consumo de energía	96	88	80	64	79	82	97	

a) Grafíquense los datos. ¿Sugiere la gráfica una asociación lineal?
b) Para un modelo lineal simple, obténgase la ecuación estimada de regresión y grafíquese sobre la gráfica de la parte *a*.

c) Interprétense los coeficientes de regresión estimados.
d) ¿Qué se recomendaría a la compañía para mejorar el modelo empírico?

13.5. Una compañía de seguros desea determinar el grado de relación que existe entre el ingreso familiar x y el monto del seguro de vida Y del jefe de familia. Con base en una muestra aleatoria de 18 familias, se obtuvo la siguiente información (en miles de dólares).

Ingreso	45	20	40	40	47	30	25	20	15
Seguro de vida	70	50	60	50	90	55	55	35	40
Ingreso	35	40	55	50	60	15	30	35	45
Seguro de vida	65	75	105	110	120	30	40	65	80

Repítanse todos los incisos del ejercicio 13.4.

13.6. Dada la ecuación de regresión estimada para el ejercicio 13.4:

a) Calcúlense los residuos.
b) Verifíquese que se cumplen las propiedades 2 y 3 de la sección 13.3.
c) Obténgase la varianza residual.
d) Calcúlense los estimadores de las desviaciones estándar de B_0 y B_1.
e) Obténgase un intervalo estimado de confianza del 95% para el valor real de la pendiente.
f) Determínese si una relación lineal entre la temperatura atmosférica promedio y el consumo de energía es estadísticamente discernible para un nivel $\alpha = 0.05$
g) Para cada temperatura atmosférica, calcúlense los intervalos de confianza del 95% estimados para el uso medio de energía y grafíquense éstos contra la recta estimada de regresión.

13.7. Repítanse todos los incisos del ejercicio 13.6 para la ecuación de regresión estimada del ejercicio 13.5.

13.8. Con respecto al ejercicio 13.4, estímense los consumos individuales de energía para las siguientes temperaturas: -10, -8, -5, -2, 1, 4, 7, 10, y 13. Obténganse intervalos de predicción del 95% para las estimaciones.

13.9. Con respecto al ejercicio 13.5, estímense los montos individuales del seguro de vida para los ingresos anuales de 18, 28, 38, 48 y 58 y obténganse intervalos de predicción del 95% para sus estimaciones.

13.10. Mediante el empleo de los datos de los ejercicios 13.4 y 13.5

a) Llévese a cabo un análisis de varianza para cada conjunto de datos y determínese si se puede rechazar la hipótesis nula de no regresión lineal a un nivel de $\alpha = 0.05$.
b) Compárense los resultados de la parte *a* con los que se obtienen en la parte *f* del ejercicio 13.6. Formúlese un comentario sobre la relación entre el valor de la estadística F, calculado aquí, con el de la estadística T determinado en la parte *f* del ejercicio 13.6.
c) Calcúlense los coeficientes de determinación y explíquese su significado. ¿Puede concluirse que las verdaderas ecuaciones de regresión entre la temperatura y el consumo de energía, o entre el ingreso anual y el monto del seguro de vida, son estrictamente lineales?

13.11. Los siguientes datos son las alturas X y los pesos Y de una muestra aleatoria de 10 empleados del sexo femenino de una gran empresa.

Altura (pulgadas)	68	67	65	68	64	67	66	65	64	66
Peso (libras)	119	118	129	135	123	140	125	132	118	130

a) Grafíquense los datos.
b) Calcúlese el coeficiente de correlación de la muestra y formúlese un comentario sobre cualquier linealidad aparente entre la altura y el peso.

13.12. Los siguientes datos* representan la potencia diaria, en megawatts, generada por una central eléctrica de servicio regional, durante el mes de agosto de 1980, y la temperatura atmosférica en grados Fahrenheit registrada a las 11 a.m. en una localidad central.

Temperatura	99	99	99	99	99	96	96	97	97
Potencia	153.4	141.0	143.1	156.8	158.7	158.5	158.7	159.6	148.3
Temperatura	97	99	94	9'	97	96	85	79	76
Potencia	137.8	160.0	154.0	142.2	149.4	147.9	114.2	94.7	112.5
Temperatura	84	90	76	78	81	90	93	90	96
Potencia	123.6	131.1	119.4	111.9	103.5	103.7	125.4	129.0	135.6
Temperatura	98	95	95	95					
Potencia	142.3	142.5	128.9	124.3					

a) Grafíquense los datos.
b) Calcúlese el coeficiente de correlación de la muestra y formúlese un comentario sobre cualquier linealidad aparente entre la temperatura y la cantidad de potencia generada.

13.13. Supóngase que se sabe que la curva de regresión entre una respuesta Y y una variable de predicción x es lineal. Para estimar la ecuación de regresión, se toman $n/2$ observaciones de Y en el extremo inferior del intervalo de valores de x y $n/2$ observaciones en el extremo superior de x. Por conveniencia, los valores extremos de x se han escalado a -1 y $+1$.

a) Empléese la ecuación (13.18) para obtener el intervalo para la varianza de la respuesta media para cualquier punto x_p de x que se encuentre dentro del intervalo $(-1, 1)$.
b) Úsese la ecuación (13.20) para obtener una expresión similar para la varianza de una respuesta en particular.
c) Supóngase que se registran las siguientes observaciones:

x	-1	-1	-1	-1	-1	1	1	1	1	1
Y	10	12	9	13	8	20	17	23	24	19

* Cortesía de K. L. Fugett.

Úsese álgebra de matrices para obtener las estimaciones de mínimos cuadrados para la pendiente y la intersección.

13.14. Supóngase que la siguiente información sobre el ingreso anual bruto x y el porcentaje de impuestos pagados Y, proviene de una muestra aleatoria de 14 declaraciones de impuestos.

Ingreso bruto (miles de dólares)	25.6	42.2	57.6	98.8	10.4	30.1	40.0
Porcentaje pagado en impuesto	15.4	16.8	19.7	21.7	10.8	15.2	18.9

Ingreso bruto (miles de dólares)	29.3	16.1	18.0	88.2	34.0	22.1	70.0
Porcentaje pagado en impuesto	15.9	12.0	14.1	21.1	17.6	14.8	21.6

a) Grafíquense los datos. ¿Sugiere esta gráfica una asociación lineal?
b) Mediante la suposición de un ajuste lineal, estímese la ecuación de regresión y dibújese la recta sobre la gráfica de la parte *a*.
c) Realícese un análisis de varianza, obténganse los coeficientes de determinación y coméntese si se puede pensar que la ecuación de regresión estimada proporciona una predicción apropiada. Úsese $\alpha = 0.05$.
d) Predígase el porcentaje promedio de impuestos que pagará la gente con ingresos brutos de 15 y 85 mil dólares y obténganse las estimaciones de sus desviaciones estándar.

13.15. El gerente de una industria desea determinar si existe una relación lineal entre el número de unidades Y, armadas por los operadores de una línea de ensamble, y el lapso x que transcurre antes de que se presente una falla. Con base en una muestra aleatoria de operadores de la línea de ensamble, se observa la siguiente información:

Tiempo (en horas)	1	2	3	4
Unidades ensambladas	25, 29, 23, 31	55, 65, 63, 59	73, 75, 74, 71	90, 88, 91, 87

a) Grafíquense los datos y coméntese el resultado.
b) Estímese una ecuación de regresión lineal mediante el uso del álgebra de matrices.
c) Determínese si la relación lineal es estadísticamente discernible para un nivel $\alpha = 0.01$
d) Obténgase un intervalo de confianza del 95% para la pendiente.

13.16. Los siguientes datos muestran el porcentaje de la población con cuatro o más años de educación superior x, y la tasa de mortalidad infantil por cada 1 000 nacimientos Y para una muestra de 15 estados.*

a) Grafíquense los datos y calcúlese el coeficiente de correlación de la muestra.
b) Ajústese una función de regresión lineal con la tasa de mortalidad como la respuesta y el porcentaje de la población con cuatro o más años de educación superior como la variable de predicción. Interprétese el coeficiente de regresión estimado para la pendiente.

* *Hammond almanac, 1981.*

x	19.4	12.3	13.7	11.0	11.5	16.8	11.8	12.8
Y	12.0	15.4	16.0	14.2	17.9	11.9	14.2	12.7
x	15.3	11.8	11.7	10.4	17.5	15.6	16.1	
Y	13.8	15.8	13.7	17.6	10.1	10.1	12.1	

c) La regresión lineal, ¿es estadísticamente discernible para un nivel $\alpha = 0.05$ ¿Cómo podría explicarse cualquier asociación lineal que existiese entre estas dos cantidades?

13.17. Los datos* que figuran en la tabla 13.8 consisten en información anual sobre los precios relativos del alcohol x y el consumo *per cápita* en litros de alcohol absoluto Y para el periodo 1948-1967 en Ontario.

a) Grafíquense los datos y calcúlese el coeficiente de correlación de la muestra.

b) Mediante el empleo del análisis de varianza, determínese si la regresión lineal entre el precio relativo y el consumo *per cápita* es estadísticamente discernible para un nivel $\alpha = 0.01$

13.18. Se llevó a cabo un estudio para determinar la relación entre el número de años de experiencia a x y el salario anual Y para una profesión en particular en una región geográfica dada. Se seleccionó una muestra aleatoria de 17 personas, las cuales ejercen esta profesión, y se obtuvo la siguiente información:

TABLA 13.8 Datos de la muestra para el ejercicio 13.17

Año	*Precio relativo*	*Consumo per cápita*
1948	0.057	7.09
1949	0.058	7.18
1950	0.055	7.23
1951	0.052	7.23
1952	0.051	7.32
1953	0.055	7.64
1954	0.056	7.73
1955	0.047	7.55
1956	0.045	7.91
1957	0.044	7.86
1958	0.043	7.96
1959	0.043	7.77
1960	0.043	8.14
1961	0.043	8.14
1962	0.041	8.23
1963	0.040	8.46
1964	0.039	8.73
1965	0.038	8.77
1966	0.039	9.18
1967	0.035	8.91

* R.E. Popham, W. Schmidt, y J. de Lint, *The prevention of alcoholism: Epidemiological studies of the effects of government control measures,* Brit. J. of Addiction 70 (1975), 125—144.

Años de experiencia	13	16	30	2	8	31	19	20	1
Salario anual actual (miles de dólares)	26.1	33.2	36.1	16.5	26.4	36.4	33.8	36.5	16.9

Años de experiencia	4	27	25	7	15	13	6	10
Salario anual actual (miles de dólares)	19.8	36.0	36.5	21.4	31.0	31.4	19.1	24.6

a) Grafíquense los datos y, con base en esta gráfica, determínese si un ajuste lineal es suficiente.

b) Ajústese un modelo lineal e interprétense los coeficientes de regresión estimados.

c) ¿Puede rechazarse la hipótesis nula de pendiente cero para un nivel $\alpha = 0.01$?

d) Estímese el salario promedio para una persona que ejerce esta profesión, la cual tiene 12 años de experiencia; además, calcúlese un intervalo de confianza del 99% para este valor.

e) Obténganse los residuos y grafíquense contra los correspondientes años de experiencia, ¿se observa algo fuera de lo común? Explíquese.

13.19. Los siguientes datos representan el producto nacional bruto x y los gastos de consumo Y en miles de millones de dólares en 1972, para los años 1960-1980.*

Año	1960	1961	1962	1963	1964	1965	1966
x	737.2	756.6	800.3	832.5	876.4	929.3	984.8
Y	452.0	461.4	482.0	500.5	528.0	557.5	585.7

Año	1967	1968	1969	1970	1971	1972	1973
x	1 011.4	1 058.1	1 087.6	1 085.6	1 122.4	1 185.9	1 255.0
Y	602.7	634.4	657.9	672.1	696.8	737.1	768.5

Año	1974	1975	1976	1977	1978	1979	1980
x	1 248.0	1 233.9	1 300.4	1 371.7	1 436.9	1 483.0	1 480.7
Y	763.6	780.2	823.7	863.9	904.8	930.9	935.1

a) Ajústese un modelo lineal e interprétense los coeficientes de regresión estimados.

b) Hágase una gráfica de los residuos estandarizados contra el tiempo. ¿Se puede detectar algún patrón?

c) Calcúlese la estadística de Durbin-Watson y determínese si los errores se encuentran positivamente autocorrelacionados. Úsese $\alpha = 0.05$.

d) Si la autocorrelación positiva es estadísticamente discernible, ajústese la ecuación de regresión estimada mediante la transformación de los datos.

13.20. Los siguientes datos representan las ganancias de las empresas por inventario y ajustes al capital x y los impuestos sobre estas ganancias Y en miles de millones de dólares para los años 1960-1980.* Repítanse todas las partes del ejercicio 13.19.

Año	1960	1961	1962	1963	1964	1965	1966
x	47.6	48.6	56.6	62.1	69.2	80.0	85.1
Y	22.7	22.8	24.0	26.2	28.0	30.9	33.7

* *Economic report of the president* february 1982.

Año	1967	1968	1969	1970	1971	1972	1973
x	82.4	89.1	85.1	71.4	83.2	96.6	108.3
Y	32.5	39.2	39.5	34.2	37.5	41.6	49.0

Año	1974	1975	1976	1977	1978	1979	1980
x	94.9	110.5	138.1	164.7	185.5	196.8	182.7
Y	51.6	50.6	63.8	72.6	83.0	87.6	82.3

APÉNDICE

Breve revisión del álgebra de matrices

Una matriz es un arreglo rectangular de elementos en renglones y columnas. Por ejemplo,

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mj} & \cdots & x_{mn} \end{bmatrix}$$

es una matriz que contiene **m** renglones y n columnas. Las entradas x_{ij}, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, son los elementos de la matriz **X**. El primer índice (i) identifica el renglón en el que se encuentra el elemento, y el segundo (j) la columna a la que pertenece. La matriz **X** de **m** renglones y n columnas se conoce como una matriz de orden (o dimensión) m por n. En general, una matriz se denota por una letra mayúscula en negritas, mientras que la correspondiente letra minúscula designa a un elemento de ésta. Es una práctica común utilizar la siguiente notación abreviada:

$$\mathbf{X} = [x_{ij}], \qquad i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots, n$$

para designar a la matriz **X** de dimensión $m \times n$.

Una matriz que contiene sólo una columna recibe el nombre *vector columna,* y una matriz formada por un renglón *vector renglón*. Las matrices

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \mathbf{Z}' = [z_1 \quad z_2 \quad \cdots \quad z_n]$$

son ejemplos de vectores columna y renglón, respectivamente; **Y** es un vector columna de $n \times 1$, y **Z**′ es un vector renglón de $1 \times n$. La razón para emplear el símbolo

de virgulilla en el vector renglón $\mathbf{Z}'$ se explicará en forma breve. Ya que un vector columna o renglón tienen sólo un renglón o una sola columna, únicamente es necesario usar una notación para identificar la posición de los elementos.

Una matriz que tiene el mismo número de renglones que de columnas recibe el nombre de *matriz cuadrada*.

$$\mathbf{A} = \begin{bmatrix} 3 & -2 & 1 \\ 0 & 2 & 4 \\ -3 & 1 & 5 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 10 & -3 \\ 2 & 1 \end{bmatrix}$$

son ejemplo de matrices cuadradas. $\mathbf{A}$ es una matriz cuadrada de 3×3, y $\mathbf{B}$ es una matriz cuadrada de 2×2.

El intercambio de los renglones y las columnas de una matriz $\mathbf{X}$ de $m \times n$ da origen a una nueva matriz denotada por $\mathbf{X}'$, de dimensión $n \times m$ que recibe el nombre de *transpuesta* de $\mathbf{X}$. Por ejemplo, dada la matriz

$$\mathbf{X} = \begin{bmatrix} 2 & -3 & 0 \\ 1 & 5 & -12 \end{bmatrix},$$

la transpuesta de $\mathbf{X}$ es la matriz cuya primera columna es igual al primer renglón de $\mathbf{X}$ y cuya segunda columna es igual al segundo renglón de $\mathbf{X}$,

$$\mathbf{X}' = \begin{bmatrix} 2 & 1 \\ -3 & 5 \\ 0 & -12 \end{bmatrix}.$$

En general, dada

$$\mathbf{X} = [x_{ij}], \qquad i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots, n,$$

se tiene

$$\mathbf{X}' = [x_{ji}], \qquad j = 1, 2, \ldots, n, \quad i = 1, 2, \ldots, m.$$

En otras palabras, el elemento en el i-ésimo renglón y la j-ésima columna de $\mathbf{X}$ se encuentra en el j-ésimo renglón y la i-ésima columna de la matriz transpuesta $\mathbf{X}'$. La transpuesta de un vector columna es un vector renglón y viceversa. Por esta razón se acostumbra emplear el símbolo de virgulilla para denotar un vector renglón.

Se dice que dos matrices son iguales sólo si sus correspondientes elementos son iguales. De esta forma, una condición necesaria para que dos matrices sean iguales es que tengan la misma dimensión. Por ejemplo, las dos matrices

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} -2 & 0 \\ 5 & 6 \\ 12 & -5 \end{bmatrix}$$

son iguales si

$$\begin{aligned} a_{11} &= -2 & a_{12} &= 0 \\ a_{21} &= 5 & a_{22} &= 6 \\ a_{31} &= 12 & a_{32} &= -5. \end{aligned}$$

La suma o diferencia entre dos matrices sólo es posible cuando sus dimensiones son las mismas. La suma (diferencia) de dos matrices es una matriz cuyos elementos son las sumas (diferencias) de los correspondientes elementos de las dos matrices. Por ejemplo, dadas

$$\mathbf{A} = \begin{bmatrix} -2 & 5 \\ 3 & 8 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 4 & -3 \\ 2 & -6 \end{bmatrix},$$

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} -2 + 4 & 5 - 3 \\ 3 + 2 & 8 - 6 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 5 & 2 \end{bmatrix},$$

$$\mathbf{A} - \mathbf{B} = \begin{bmatrix} -2 - 4 & 5 - (-3) \\ 3 - 2 & 8 - (-6) \end{bmatrix} = \begin{bmatrix} -6 & 8 \\ 1 & 14 \end{bmatrix}.$$

Dadas dos matrices **A** y **B**, la matriz producto **AB** se define sólo si el número de columnas de **A** es igual al número de renglones de **B**. Entonces, si **A** es de $m \times n$ y **B** es de $n \times p$, el producto **AB** es una matriz de dimensión $m \times p$ para la que el elemento que se encuentra en el i-ésimo renglón y la j-ésima columna es igual a la suma de los productos de los elementos que se encuentran en el i-ésimo renglón de **A** y la j-ésima columna de **B**. Si

$$\mathbf{A} = \begin{bmatrix} 1 & -2 \\ -3 & 4 \\ 0 & -1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -2 & 1 \\ 4 & 3 \end{bmatrix},$$

$$\mathbf{AB} = \begin{bmatrix} 1 & -2 \\ -3 & 4 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} -2 & 1 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} (1)(-2) + (-2)(4) & (1)(1) + (-2)(3) \\ (-3)(-2) + \quad (4)(4) & (-3)(1) + \quad (4)(3) \\ (0)(-2) + (-1)(4) & (0)(1) + (-1)(3) \end{bmatrix}$$

$$= \begin{bmatrix} -10 & -5 \\ 22 & 9 \\ -4 & -3 \end{bmatrix}.$$

Nótese que para este par de matrices el producto **BA** no está definido; en general, la multiplicación de matrices no es conmutativa. También es interesante notar que si **Y** es un vector columna de $n \times 1$ y **Y**′ es un vector renglón de $1 \times n$, entonces **YY**′ es una matriz cuadrada de dimensión n y **Y**′**Y** es un escalar. Un *escalar* es cualquier número de la recta real $(-\infty, \infty)$. La multiplicación de una matriz por un escalar da origen a una matriz cuyos elementos son los productos de los correspondientes elementos originales y la cantidad escalar. Por ejemplo, dada

$$\mathbf{A} = \begin{bmatrix} -2 & 1 & -2 \\ 3 & 4 & 1 \end{bmatrix},$$

$$-5\mathbf{A} = \begin{bmatrix} (-5)(-2) & (-5)(1) & (-5)(-2) \\ (-5)(3) & (-5)(4) & (-5)(1) \end{bmatrix} = \begin{bmatrix} 10 & -5 & 10 \\ -15 & -20 & -5 \end{bmatrix}.$$

Existen ciertas matrices especiales que vale la pena mencionar. Una matriz cuadrada de dimensión n cuyos elementos son cero excepto los que se encuentran sobre la diagonal principal,* elementos iguales a uno, recibe el nombre de *matriz identidad* de orden n. Por ejemplo,

$$\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

son matrices identidad de orden 3 y 2, respectivamente. En la multiplicación de matrices cuadradas, la matriz identidad juega el mismo papel que el número 1 tiene en la multiplicación entre escalares. Esto es, dada cualquier matriz **A**, el producto de la correspondiente matriz identidad y **A** da como resultado la matriz **A**, siempre que exista compatibilidad para llevar a cabo la multiplicación. De esta forma,

$$\mathbf{IA} = \mathbf{AI} = \mathbf{A}.$$

Se dice que una matriz cuadrada es *simétrica,* si es igual a su transpuesta. Dada cualquier matriz cuadrada **A**, si $\mathbf{A} = \mathbf{A}'$, entonces **A** es simétrica. Por ejemplo,

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & -2 \\ 1 & 4 & 3 \\ -2 & 3 & 1 \end{bmatrix}$$

es una matriz simétrica. Nótese que los elementos que se encuentran formando un triángulo por debajo de la diagonal principal son idénticos a los correspondientes en el triángulo que se encuentra por encima de la diagonal principal. Si una matriz **A** de $m \times n$ se premultiplica por su transpuesta, la matriz producto será simétrica. De esta forma, $\mathbf{A}'\mathbf{A}$ es una matriz simétrica de orden n. Por ejemplo, dadas

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 4 \\ 3 & 2 \end{bmatrix}, \qquad \mathbf{A}' = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 4 & 2 \end{bmatrix},$$

$$\mathbf{A}'\mathbf{A} = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 4 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 4 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 14 & 12 \\ 12 & 21 \end{bmatrix}.$$

Una *matriz diagonal* es cualquier matriz cuadrada para la que todos los elementos que se encuentran fuera de la diagonal principal son cero.

$$\mathbf{A} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$

es una matriz diagonal. Debe ser evidente que la matriz identidad es un caso especial de una matriz diagonal.

* La diagonal principal contiene los elementos cuyas posiciones en el renglón y la columna son las mismas.

Un vector cero es cualquier vector columna para el que todos sus elementos son cero.

$$\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

es un vector cero de 4 × 1.

A continuación se define un concepto importante en el álgebra matricial, que se conoce como la inversa de una matriz cuadrada. Sea **A** una matriz cuadrada de orden *n*. Si existe una matriz denotada por $\mathbf{A}^{-1}$, tal que

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$

donde **I** es la correspondiente matriz identidad, entonces $\mathbf{A}^{-1}$ es la ***única matriz inversa de*** **A**. Si una matriz cuadrada tiene inversa, se dice que es *no singular*; de otra forma, recibe el nombre de matriz *singular*.

Para cada matriz cuadrada, es posible definir y calcular una cantidad escalar que se conoce como el *determinante* de la matriz. El valor del determinante de una matriz cuadrada es el factor para decidir si ésta tiene o no inversa. Sea **A** cualquier matriz cuadrada. Si el determinante de **A**, denotado por det(**A**) no es igual a cero, existe la matriz inversa de **A**. Si det(**A**) = 0, entonces **A** es singular. La noción de una matriz inversa es el análogo del inverso multiplicativo en el álgebra de escalares.

Como ilustración, se encontrará la inversa de matrices sólo para el caso de 2 × 2. En general, sea

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

cualquier matriz de 2 × 2. El determinante de **A** se define como

$$\det(\mathbf{A}) = a_{11}a_{22} - a_{12}a_{21}$$

y puede demostrarse que la matriz inversa de **A** está dada por

$$\mathbf{A}^{-1} = \begin{bmatrix} \dfrac{a_{22}}{\det(\mathbf{A})} & -\dfrac{a_{12}}{\det(\mathbf{A})} \\ -\dfrac{a_{21}}{\det(\mathbf{A})} & \dfrac{a_{11}}{\det(\mathbf{A})} \end{bmatrix}$$

Por ejemplo, dada

$$\mathbf{A} = \begin{bmatrix} 2 & 3 \\ -1 & 1 \end{bmatrix},$$

$$\det(\mathbf{A}) = (2)(1) - (-1)(3) = 5,$$

y

$$\mathbf{A}^{-1} = \begin{bmatrix} 1/5 & -3/5 \\ 1/5 & 2/5 \end{bmatrix}$$

es la matriz inversa de **A**. Este resultado puede verificarse en forma sencilla, ya que

$$\mathbf{A}^{-1}\mathbf{A} = \begin{bmatrix} 1/5 & -3/5 \\ 1/5 & 2/5 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ -1 & 1 \end{bmatrix} = \mathbf{A}\mathbf{A}^{-1}$$

$$= \begin{bmatrix} 2 & 3 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1/5 & -3/5 \\ 1/5 & 2/5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Para finalizar, debe notarse que la inversa de cualquier matriz diagonal también es una matriz diagonal, cuyos elementos sobre la diagonal principal son los recíprocos de los elementos que se encuentran en la diagonal principal de la matriz original. Por ejemplo. dado que

$$\mathbf{A} = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 10 \end{bmatrix},$$

$$\mathbf{A}^{-1} = \begin{bmatrix} 1/9 & 0 & 0 \\ 0 & 1/5 & 0 \\ 0 & 0 & 1/10 \end{bmatrix}.$$

CAPÍTULO CATORCE

Análisis de regresión: el modelo lineal general

14.1 Introducción

En el capítulo 13 se examinaron los fundamentos del análisis de regresión para el modelo lineal simple. En este capítulo se extenderán los conceptos ya presentados al modelo lineal general para el cual una respuesta dada se considera como una función de varias variables de predicción. Al examinar este modelo se estudiarán algunas formas para determinar el mejor conjunto de variables de predicción por incluir en la ecuación de regresión. También se proporcionará un estudio detallado del análisis de residuos (también conocidos como residuales), mínimos cuadrados con factores de peso (ponderados) y variables indicadoras, así como ejemplos resueltos con gran detalle. Para este capítulo se emplearán los paquetes estadísticos para computadoras Minitab y SAS (véase [6]). Se supone que este tipo de paquetes o algunos similares se encuentran disponibles para el lector. Para un estudio más teórico de los temas presentados en este capítulo, se invita al lector a que consulte [4].

14.2 El modelo lineal general

Sean $x_1, x_2, \ldots, x_k$ k variables de predicción, las cuales pueden tener alguna influencia sobre una respuesta Y, y supóngase que el modelo tiene la forma donde Y_i es la

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i, \qquad i = 1, 2, \ldots, n, \tag{14.1}$$

i-ésima observación de la respuesta para un conjunto de valores fijos $x_{i1}, x_{i2}, \ldots, x_{ik}$ de las variables de predicción, ε_i es el error aleatorio no observable asociado con Y_i, y $\beta_0, \beta_1, \ldots, \beta_k$ son $m = k + 1$ parámetros lineales desconocidos. La ecuación (14.1) recibe el nombre de *modelo lineal general* y da origen a lo que se conoce como una *regresión lineal múltiple*.

Si se supone el caso de la teoría basada en el modelo normal, las observaciones Y_i son variables aleatorias independientes, normalmente distribuidas con

$$E(Y_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik},$$

$$Var(Y_i) = \sigma^2, \qquad i = 1, 2, \ldots, n.$$

De esta forma, los errores aleatorios ε_i son $N(0, \sigma^2)$ independientes. El modelo lineal general define una ecuación de regresión la cual representa un hiperplano, para la que el parámetro β_0 es el valor de la respuesta media cuando todas las variables de predicción tienen un valor igual a cero. El parámetro β_j, $j = 1, 2, \ldots, k$, representa el cambio en la respuesta promedio para un cambio igual a una unidad de la correspondiente variable de predicción x_j, cuando todas las demás variables de predicción se mantienen constantes. En este sentido, β_j representa el efecto parcial de x_j sobre la respuesta.

La única restricción funcional que se impone al modelo lineal general es que sea lineal en los parámetros desconocidos; el modelo no tiene ninguna restricción con respecto a la naturaleza de las variables de predicción; por lo tanto surgen muchos casos especiales e interesantes, algunos de los cuales cabe mencionar. El modelo dado por (14.1) implica que los efectos que las variables de predición $x_1, x_2, \ldots, x_k$ tienen sobre la respuesta son aditivos, de tal manera que la ecuación de regresión propuesta es una función lineal de las variables de predicción. Una ecuación de este tipo se denomina *modelo de primer orden.* Sin embargo, es posible que dos o más variables de predicción interactúen, es decir, el efecto de una de las variables de predicción sobre la variable de respuesta depende del valor de otra variable de predicción. Cuando esto ocurre, los efectos no son aditivos debido a la presencia en el modelo de un término que contiene un producto cruzado el cual representa el efecto de interacción. Por ejemplo, considérese un modelo que contiene dos variables de predicción que interactúan. El modelo es

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \varepsilon_i, \tag{14.2}$$

donde el sumando $\beta_3 x_{i1} x_{i2}$ refleja la interacción entre x_1 y x_2. Si se define

$$x_{i3} = x_{i1} x_{i2}, \qquad i = 1, 2, \ldots, n,$$

entonces (14.2) puede escribirse en la forma del modelo lineal general (14.1), y de esta manera se advierte que es un caso especial de éste. Nótese que para este caso especial el significado de β_1 y β_2 no es el mismo dado con anterioridad. La derivada parcial de la respuesta media con respecto a x_1 (o con respecto a x_2) representa el efecto sobre la respuesta media por unidad de cambio en x_1 (x_2) cuando x_2 (x_1) se mantiene fija. Las derivadas parciales son

$$\frac{\partial E(Y)}{\partial x_1} = \beta_1 + \beta_3 x_2$$

$$\frac{\partial E(Y)}{\partial x_2} = \beta_2 + \beta_3 x_1.$$

Otro caso interesante surge cuando en (14.1) se tiene

$$x_{ij} = x_i^j, \qquad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, k.$$

Entonces el modelo lineal general toma la forma

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_k x_i^k + \varepsilon_i, \tag{14.3}$$

la cual se conoce como *modelo curvilineal o polinomial.* En este caso se supone que la respuesta promedio es una función polinómica de grado k de una sola variable de predicción. Por lo tanto, la ecuación de regresión propuesta para la respuesta promedio es una función no lineal de la variable de predicción, pero sigue siendo lineal en los parámetros. Es importante notar que lo que se busca en este caso es el grado k que mejor se ajusta a una muestra aleatoria de la variable respuesta.

Para describir en forma adecuada una variable respuesta dada, muchas veces es necesario incluir términos lineales, cuadráticos y de interacción en el modelo propuesto. Por ejemplo, un modelo para dos variables de predicción podría ser

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1}^2 + \beta_4 x_{i2}^2 + \beta_5 x_{i1} x_{i2} + \varepsilon_i. \tag{14.4}$$

Al definir nuevas variables de predicción, así como se hizo anteriormente para los términos cuadráticos y de interacción, se observa que (14.4) también es un caso especial del modelo general. Este tipo de modelo se denomina *ecuación completa de segundo orden* y define varias superficies para la respuesta promedio como una función no lineal de las variables de predicción x_1 y x_2. Para $k \geqslant 2$ variables de predicción distintas, una ecuación de regresión completa de segundo orden consiste en un término constante, k términos lineales, k términos cuadráticos y $k(k-1)/2$ términos de interacción.

A continuación se regresará al modelo lineal general dado en (14.1) para obtener los estimadores de mínimos cuadrados de los parámetros y para desarrollar técnicas de regresión para este modelo. Todos los casos especiales mencionados con anterioridad así como muchos otros que no se citaron de manera específica, se encuentran incluidos en el siguiente análisis. Se empleará el álgebra de matrices, ya que ésta simplifica en gran medida la presentación.

Dada una muestra aleatoria de observaciones $Y_1, Y_2, \ldots, Y_n$ en los puntos de observación $x_{11}, x_1\, x_{12}, \ldots, x_{1k}, x_{21}, x_{22}, \ldots, x_{2k}, \ldots, x_{n1}, x_{n2}, \ldots, x_{nk}$, respectivamente, con base en el modelo lineal general, se tienen las n ecuaciones siguientes:

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_k x_{2k} + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_k x_{nk} + \varepsilon_n. \end{aligned}$$

Como resultado, el modelo lineal general también puede expresarse en forma matricial como

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{14.5}$$

donde

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

El lector no tendrá ninguna dificultad para reconocer que (14.5) tiene la misma forma matricial que el modelo lineal simple (13.37), excepto que ahora **X** es una matriz de $n \times m$ para las variables de predicción, y $\boldsymbol{\beta}$ es un vector de parámetros desconocidos de $m \times 1$, mientras que **Y** y $\boldsymbol{\varepsilon}$ siguen siendo vectores de $n \times 1$, los que contienen las observaciones de la variable de respuesta y los errores aleatorios asociados con éstas, respectivamente.

Bajo el caso de la teoría normal

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}),$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}),$$

donde

$$Var(\mathbf{Y}) = Var(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}.$$

De esta manera **Y** y $\boldsymbol{\varepsilon}$ son vectores de variables aleatorias independientes normalmente distribuidas.

Para la estimación de los parámetros por mínimos cuadrados las ecuaciones normales toman la misma forma dada por (13.40), o

$$(\mathbf{X}'\mathbf{X})\mathbf{B} = \mathbf{X}'\mathbf{Y}$$

donde, ahora, $(\mathbf{X}'\mathbf{X})$ es un matriz de $m \times n$ y **B** es un vector de $m \times 1$ el cual contiene los estimadores de mínimos cuadrados $B_0, B_1, \ldots, B_k$. Si $(\mathbf{X}'\mathbf{X})$ tiene inversa, la solución para el vector **B** está dada por

$$\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

Por lo tanto, la ecuación estimada de regresión es

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{B}, \tag{14.6}$$

donde el vector $\hat{\mathbf{Y}}$ de $n \times 1$ contiene los valores estimados para la respuesta promedio correspondientes a los n puntos de observación de las variables de predicción. La diferencia entre los vectores **Y** y $\hat{\mathbf{Y}}$ proporciona el vector de residuos.

Puede demostrarse que las propiedades de los estimadores de mínimos cuadrados $B_0, B_1, \ldots, B_k$ son extensiones de las propiedades de los estimadores para el modelo lineal simple, es decir, de acuerdo con el caso de la teoría normal, los estimadores también son de máxima verosimilitud, de tal manera que lo siguiente se verifica:

1. Cada B_j tiene una distribución normal con media $E(B_j) = \beta_j$, $j = 0, 1, 2, \ldots, k$, y varianza $Var(B_j) = c_{(j+1)}\sigma^2$, $j = 0, 1, 2, \ldots, k$, donde $c_{(j+1)}$ es el elemento de la diagonal $(j + 1)$ de $(\mathbf{X}'\mathbf{X})^{-1}$.
2. $Cov(B_i, B_j) = c_{(i+1),(j+1)}\sigma^2$, $i \neq j = 0, 1, 2, \ldots, k$, donde $c_{(i+1),(j+1)}$ es el elemento de $(\mathbf{X}'\mathbf{X})^{-1}$ que se encuentra en el renglón $(i + 1)$ y la columna $(j + 1)$ para $i \neq j$.

Un estimador no sesgado de la varianza del error es

$$S^2 = \frac{\mathbf{Y}'\mathbf{Y} - \mathbf{B}'\mathbf{X}'\mathbf{Y}}{n - m}, \tag{14.7}$$

donde el numerador de (14.7) no es más que la suma de los cuadrados de los residuos. Nótese que el denominador de (14.7) es igual al número de observaciones, menos el número de parámetros que figuran en el modelo, el que para el modelo lineal general es $m = k + 1$. Por lo tanto, una estimación de $Var(B_j)$ es

$$s^2(B_j) = c_{(j+1)}s^2, \qquad j = 0, 1, 2, \ldots, k,$$

donde $c_{(j+1)}$ tiene un valor igual al ya definido con anterioridad.

De los resultados anteriores puede deducirse que la cantidad

$$(B_j - \beta_j)/s(B_j), \qquad j = 0, 1, 2, \ldots, k,$$

es una variable aleatoria t de Student con $n - m$ grados de libertad. Entonces, un intervalo de confianza del $100(1 - \alpha)\%$ para el parámetro β_j es

$$b_j \pm t_{1-\alpha/2,\, n-m}s(B_j), \qquad j = 0, 1, 2, \ldots, k, \tag{14.8}$$

y una estadística apropiada para probar la hipótesis nula

$$H_0\colon \beta_j = 0$$

contra cualquier alternativa, ya sea ésta uni o bilateral, es la ya familiar t de Student

$$T = B_j/s(B_j), \qquad j = 0, 1, 2, \ldots, k,$$

con $n - m$ grados de libertad.

Considérese la técnica del análisis de varianza para probar la hipótesis nula

$$H_0\colon \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

contra la alternativa

$$H_1\colon \beta_j \neq 0 \text{ para algún } j = 1, 2, \ldots, k.$$

Dado que H_0 establece que todos los parámetros de regresión son iguales a cero, excepto el término constante, esto implica que no existe ninguna relación igual a la especificada por el modelo propuesto entre la respuesta y el conjunto de variables de predicción. No obstante, se advierte al lector que el hecho de rechazar a H_0 no nece-

sariamente implica que la ecuación estimada de regresión sea útil para efectuar predicciones. Se necesita profundizar el análisis antes de que se pueda dar un juicio definitivo sobre la utilidad de la ecuación de regresión.

Al seguir el mismo argumento para el modelo lineal general que el dado para el modelo lineal simple en la sección 13.7, puede demostrarse que la suma total de cuadrados se encuentra dividida en la suma de cuadrados de la regresión y en la suma de cuadrados de los errores. Mediante el empleo de la notación matricial, *STC, SCR* y *SCE* se encuentran definidos en la tabla 14.1.

El número total de grados de libertad sigue siendo $n - 1$, pero el número de grados de libertad para el error ahora es de $n - m$. Los grados de libertad para la regresión son $(n - 1) - (n - \mathrm{m}) = m - 1 = k$, dado que $m = k + 1$. La varianza residual o $SCE/(n - m)$ es el cuadrado medio del error y $SCR/(m - 1)$ es el cuadrado medio de la regresión. Bajo la hipótesis nula, la estadística de prueba apropiada es

$$F = \mathrm{CMR/CME},$$

la cual tiene una distribución F con $m - 1$ y $n - m$ grados de libertad. Al igual que en los casos anteriores, puede argumentarse que si un valor de esta estadística es lo suficientemente grande, entonces una porción considerable de la variación en las observaciones puede atribuirse a la regresión de Y sobre las variables de predicción como se encuentran definidas por el modelo. De esta forma se rechaza la hipótesis nula siempre que el valor calculado se encuentre en el interior de una región crítica de tamaño α en el extremo superior de la distribución. En la tabla 14.1 se da la tabla de análisis de varianza para el modelo lineal general.

Para el modelo lineal general la noción del coeficiente de determinación se extiende para dar origen a lo que se conoce como *coeficiente de correlación múltiple* o *coeficiente de determinación múltiple.* El coeficiente de correlación múltiple se define como

$$R^2 = \frac{\mathrm{SCR}}{\mathrm{STC}} = 1 - \frac{\mathrm{SCE}}{\mathrm{STC}}, \tag{14.9}$$

y al igual que r^2, mide la proporción de la variación total de las observaciones con respecto a su media, atribuible a la ecuación de regresión estimada. En otras pala-

TABLA 14.1 Tabla ANOVA para el modelo lineal general

Fuente de variación	*Número de grados de libertad*	*Sumas de los cuadrados*	*Cuadrados medios*	*Estadística F*
Regresión	$k = m - 1$	$\mathbf{B'X'Y} - \frac{(\Sigma Y_i)^2}{n}$	$\mathrm{SCR}/(m - 1)$	$\frac{\mathrm{SCR}/(m - 1)}{\mathrm{SCE}/(n - m)}$
Error	$n - m$	$\mathbf{Y'Y} - \mathbf{B'X'Y}$	$\mathrm{SCE}/(n - m)$	
Total	$n - 1$	$\mathbf{Y'Y} - \frac{(\Sigma Y_i)^2}{n}$		

bras, R^2 es una medida relativa de qué tanto las variables de predicción incluidas en el modelo explican la variación de las observaciones. Al igual que para el modelo lineal simple, $0 \leqslant R^2 \leqslant 1$, y entre más cercano a uno es el valor de R^2 mayor es la cantidad de la variación total que puede explicarse por medio de los términos que aparecen en el modelo. Por sí mismo, R^2 no puede validar el modelo propuesto, ni tener un valor de R^2 cercano a uno necesariamente implica que la ecuación de regresión estimada sea apropiada para predicción.

Supóngase que se desea predecir la respuesta promedio cuando las k variables de predicción toman los valores específicos $x_1, x_2, \ldots, x_k$, respectivamente. En notación matricial, sea

$$\mathbf{X}_p' = [1\ x_1\ x_2 \cdots x_k]$$

un vector renglón el cual identifica las coordenadas para las cuales se va a formular la predicción. Entonces la respuesta promedio estimada es

$$\begin{aligned} \hat{Y}_p &= \mathbf{X}_p'\mathbf{B} \\ &= B_0 + B_1x_1 + B_2x_2 + \cdots + B_kx_k. \end{aligned} \tag{14.10}$$

Dada (14.10), puede demostrarse que

$$Var(\hat{Y}_p) = \sigma^2\mathbf{X}_p'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_p.$$

De esta forma, una estimación de $Var(\hat{Y}_p)$ es

$$s^2(\hat{Y}_p) = s^2\mathbf{X}_p'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_p, \tag{14.11}$$

donde s^2 es la varianza residual y $\mathbf{X}$ es la matriz original de valores x, los cuales dieron origen a la ecuación de regresión estimada. De acuerdo con el caso de la teoría normal, un intervalo de confianza del $100(1 - \alpha)\%$ para la respuesta promedio en $x_1, x_2, \ldots, x_k$, es

$$\hat{y}_p \pm t_{1-\alpha/2,\, n-m}\, s(\hat{Y}_p). \tag{14.12}$$

Si se desea estimar una respuesta particular para $x_1, x_2, \ldots, x_k$, la predicción estará dada por (14.10), pero la varianza será

$$Var(\hat{Y}_{\text{part}}) = \sigma^2[1 + \mathbf{X}_p'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_p].$$

Por lo tanto, un intervalo de predicción del $100(1 - \alpha)\%$ para el valor real de la respuesta en $x_1, x_2, \ldots, x_k$, es

$$\hat{y}_{\text{part}} \pm t_{1-\alpha/2,\, n-m}\, s(\hat{Y}_{\text{part}}), \tag{14.13}$$

donde

$$s^2(\hat{Y}_{\text{part}}) = s^2[1 + \mathbf{X}_p'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_p].$$

Ejemplo 14.1 N.H.* Prater desarrolló una ecuación de regresión para estimar la producción de gasolina como una función de las propiedades de destilación de cierto tipo de petróleo crudo. Se identificaron cuatro variables de predicción: la gravedad del petróleo crudo, °API (x_1); la presión de vapor del petróleo crudo, psi (x_2); el punto de 10% *ASTM* para el petróleo crudo, °F (x_3) y el punto final *ASTM* para la gasolina, °F (x_4). Los primeros dos miden la gravedad y la presión de vapor del petróleo crudo. El punto de 10% *ASTM* es la temperatura para la cual se ha evaporado cierta cantidad de líquido, y el punto final para la gasolina es la temperatura para la cual se ha evaporado todo el líquido. La variable respuesta fue la cantidad de gasolina producida expresada como un porcentaje respecto al total de petróleo crudo. El objetivo radicó en determinar una ecuación de regresión para la producción de gasolina como una función lineal de las propiedades de destilación de cierto tipo de petróleo crudo x_1, x_2, x_3 y el punto final deseado para la gasolina x_4. Los datos de laboratorio obtenidos por Prater se muestran en la tabla 14.2.

Se emplearán los datos que aparecen en la tabla 14.2 para ilustrar las técnicas que hasta este momento se han presentado para regresión lineal múltiple mediante el empleo del paquete *SAS*. Este problema también se considerará como perteneciente a un problema particular que puede encontrarse en la regresión lineal múltiple y que se conoce como *multicolinealidad*. Debe notarse que desde la publicación de los datos de Prater, en 1956, varios autores los han empleado con el propósito de ilustrar diferentes aspectos del modelo lineal general. Entre ellos, Daniel y Wood [2] desarrollaron una ecuación de regresión muy diferente a la dada por Prater.

Mediante el empleo de una opción de *SAS*, denominada *GLM*, se ajusta el modelo lineal

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon.$$

En la figura 14.1 se proporciona el listado de computadora. Nótese que en la parte inferior de éste se encuentran cinco columnas de información. La primera columna de la izquierda identifica a las variables de predicción en el modelo que incluyen al término constante. La segunda columna proporciona las estimaciones por mínimos cuadrados; en la tercera se encuentran los valores t de Student para probar la hipótesis nula de que el valor del parámetro es cero; la cuarta columna da la probabilidad (valor p) de observar un valor t de Student, al menos tan grande en magnitud, como el valor observado (ignorando su signo) y la quinta columna proporciona las desviaciones estándar (errores) para las estimaciones por mínimos cuadrados. De esta forma, la ecuación estimada de regresión (tomando en cuenta sólo dos cifras decimales) es

$$\hat{y} = -6.82 + 0.23x_1 + 0.55x_2 - 0.15x_3 + 0.15x_4.$$

En la parte superior de la figura se encuentra la tabla ANOVA con $gl(CMR) = 4$, $SCR = 3\,429.27$, CMR = 857.32, $gl(SCE) = 27$, $SCE = 134.80$, $ECM = 4.99$,

* N.H. Prater, *Estimate gasoline yields from crudes*, Petroleum Refiner 35 (1956), 236-238. La reproducción de la tabla se hizo con el permiso de Petroleum Refiner (posteriormente Hydrocarbon Processing), Mayo 1956.

TABLA 14.2 Datos de la muestra para el ejemplo 14.1

Observación	Y	x_1	x_2	x_3	x_4
1	6.9	38.4	6.1	220	235
2	14.4	40.3	4.8	231	307
3	7.4	40.0	6.1	217	212
4	8.5	31.8	0.2	316	365
5	8.0	40.8	3.5	210	218
6	2.8	41.3	1.8	267	235
7	5.0	38.1	1.2	274	285
8	12.2	50.8	8.6	190	205
9	10.0	32.2	5.2	236	267
10	15.2	38.4	6.1	220	300
11	26.8	40.3	4.8	231	367
12	14.0	32.2	2.4	284	351
13	14.7	31.8	0.2	316	379
14	6.4	41.3	1.8	267	275
15	17.6	38.1	1.2	274	365
16	22.3	50.8	8.6	190	275
17	24.8	32.2	5.2	236	360
18	26.0	38.4	6.1	220	365
19	34.9	40.3	4.8	231	395
20	18.2	40.0	6.1	217	272
21	23.2	32.2	2.4	284	424
22	18.0	31.8	0.2	316	428
23	13.1	40.8	3.5	210	273
24	16.1	41.3	1.8	267	358
25	32.1	38.1	1.2	274	444
26	34.7	50.8	8.6	190	345
27	31.7	32.2	5.2	236	402
28	33.6	38.4	6.1	220	410
29	30.4	40.0	6.1	217	340
30	26.6	40.8	3.5	210	347
31	27.8	41.3	1.8	267	416
32	45.7	50.8	8.6	190	407

$gl(STC) = 31$ y $STC = 3\,564.08$. El valor F calculado para probar la hipótesis nula

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

es de 171.71, y la probabilidad de observar un valor mayor se encuentra inmediatamente a la derecha de éste. Debajo del valor p está la desviación estándar residual, $s = 2.23$. El coeficiente de correlación múltiple es 0.9622, lo cual significa que alrededor de un 96% de la variación total de las observaciones con respecto a su media puede explicarse por las cuatro variables de predicción incluidos en la ecuación de regresión.

En el extremo superior derecho, está el coeficiente de variación, el cual se definió en el capítulo 3. En este caso, el valor de CV es el cociente de la desviación estándar residual entre la media de las observaciones. Ya que en este caso $s = 2.23$ y $\bar{y} =$

VARIABLE DEPENDIENTE: Y

FUENTE	DF	SUMA DE CUADRADO	CUADRADO MEDIO	VALOR F	PR > F	R-CUADRADA	C.V.
MODELO	4	3429.27322460	857.31830615	171.71	0.0001	0.962177	11.3658
ERROR	27	134.80396290	4.99273937		STD DEV		Y MEDIA
TOTAL CORREGIDO	31	3564.07718750			2.23444386		19.65937500

FUENTE	DF	SC TIPO I	VALOR F	PR > F	DF	SC TIPO IV	VALOR F	PR > F
X1	1	216.25576661	43.31	0.0001	1	25.81557060	5.17	0.0311
X2	1	309.85082754	62.06	0.0001	1	11.19716648	2.24	0.1458
X3	1	29.21431690	5.85	0.0226	1	130.67556799	26.17	0.0001
X4	1	2873.95231355	575.63	0.0001	1	2873.95231355	575.63	0.0001

PARAMETRO	ESTIMACION	T PARA HO: PARAMETRO = 0	PR > |T|	DEST ERROR DE LA ESTIMACION
INTERSECCION	−6.82077407	−0.67	0.5062	10.12315182
X1	0.22724595	2.27	0.0311	0.09993664
X2	0.55372621	1.50	0.1458	0.36975194
X3	−0.14953562	−5.12	0.0001	0.02922920
X4	0.15465009	23.99	0.0001	0.00644584

FIGURA 14.1 Listado de computadora para la regresión lineal de Y sobre x_1, x_2, x_3 y x_4 para los datos de Prater

19.66, $CV = 11.37\%$. En el análisis de regresión es deseable que la desviación estándar residual sea una pequeña fracción de la media de las observaciones, ya que lo anterior, en general, implica que gran parte de la variación en la respuesta se explica mediante las variables de predicción en la ecuación de regresión. En la siguiente sección se darán más explicaciones con respecto a la información que se encuentra en la parte media de la figura.

Con base en el análisis anterior, existe una pequeña duda de que la regresión entre la producción de gasolina y las cuatro variables de predicción sea estadísticamente significativa. Debido a que se rechaza la hipótesis nula de que todos los coeficientes de regresión (excepto el término constante) son iguales a cero y el valor del coeficiente de correlación múltiple es relativamente alto al 0.9622. Sin embargo, existe una razón para preocuparse con respecto a la utilidad de la ecuación de regresión dada. Por ejemplo, las desviaciones estándar de los estimadores de mínimos cuadrados para β_0 y β_2 son grandes, lo que sugiere que x_2, y posiblemente otras variables de predicción, puedan no tener un gran efecto sobre la producción de gasolina. En las siguientes secciones se examinarán los procedimientos adecuados para obtener la mejor ecuación de regresión para un conjunto dado de variables de predicción. Los datos del ejemplo 14.1 se utilizarán de vez en cuando para otros ejemplos en este capítulo.

14.3 Principio de la suma de cuadrados extra

La inclusión de una variable de predicción en un modelo de regresión no implica, en forma necesaria, que tenga un efecto substancial sobre la respuesta dada; es decir, cuando un investigador identifica un conjunto de variables de predicción, esto indica el *potencial* de las variables para explicar la variación en la respuesta. Queda por comprobarse si algunas realmente lo hacen.

El procedimiento apropiado para encontrar los efectos individuales de las variables de predicción se basa en el *principio de la suma de cuadrados extra.* Este principio permite determinar la reducción en la suma de los cuadrados de los errores cuando se introduce un coeficiente adicional de regresión para alguna función de una variable de predicción en la ecuación de regresión. Cabe recordar dos cosas importantes: 1) la suma total de cuadrados sigue siendo la misma sin importar el número de términos que se introduzcan en el modelo de regresión. 2) La suma de los cuadrados de los errores siempre disminuye (cuando menos un poco) conforme se añaden más términos al modelo.

Dado que la suma de los cuadrados de regresión es la diferencia entre *STC* y *SCE,* el incremento en *SCR* tiene un límite conforme se suman más términos al modelo. Una estrategia lógica en la regresión lineal múltiple es la de añadir no cualesquiera términos, al modelo, sino sólo aquéllos que incrementen en forma significativa la suma de los cuadrados de regresión y de esta manera disminuyan significativamente la suma de los cuadrados de los errores. Como ejemplo, en el modelo lineal simple, *SCR* es la suma extra de los cuadrados debida a la inclusión del término lineal $\beta_1 x$ en el modelo. En otras palabras, *SCR* representa la reducción en la suma de los cuadrados de

los errores cuando se añade un efecto lineal de la variable de predicción al modelo original.

$$Y_i = \beta_0 + \varepsilon_i.$$

Para ilustrar el principio de la suma de cuadrados extra, se emplearán, de los datos de Prater como variables de predicción potenciales, sólo a x_2 y x_3 y se ajustarán todas las posibles regresiones de la producción de gasolina para esas dos variables. Existen tres ecuaciones de regresión; dos que toman en cuenta a x_2 y x_3 en forma individual y la tercera que contiene a ambas variables x_2 y x_3. En la tabla 14.3 se proporcionan las ecuaciones de regresión estimadas y sus correspondientes tablas de análisis de varianza. Nótese que se ha empleado la notación $SCR(x_2)$, $SCR(x_2, x_3)$ y $SCE(x_2, x_3)$, para denotar que estas sumas de cuadrados son funciones de las variables de predicción ya indicadas en la ecuación de regresión y de los correspondientes coeficientes de mínimos cuadrados.

A continuación se examinarán los resultados que se encuentran en la tabla 14.3. Como ya se ha mencionado, para las 32 observaciones dadas de la respuesta, la

TABLA 14.3 Ecuaciones estimadas de regresión y tablas ANOVA para la producción de gasolina, tomando en cuenta a x_2 y/o x_3

a) $\hat{y} = 13.09 + 1.57x_2$

Fuente de variación	*gl*	*SC*	*CM*	*Valor F*
Regresión	1	$SCR(x_2) = 525.74$	$CMR(x_2) = 525.74$	5.19
Error	30	$SCE(x_2) = 3038.34$	$CME(x_2) = 101.28$	
Total	31	$STC = 3564.08$	$f_{0.95,1,30} = 4.17$	

b) $\hat{y} = 41.39 - 0.09x_3$

Fuente de variación	*gl*	*SC*	*CM*	*Valor F*
Regresión	1	$SCR(x_3) = 353.70$	$CMR(x_3) = 353.70$	3.31
Error	30	$SCE(x_3) = 3210.38$	$CME(x_3) = 107.01$	
Total	31	$STC = 3564.08$	$f_{0.95,1,30} = 4.17$	

c) $\hat{y} = -2.52 + 2.26x_2 + 0.05x_3$

Fuente de variación	*gl*	*SC*	*CM*	*Valor F*
Regresión	2	$SCR(x_2, x_3) = 547.49$	$CMR(x_2, x_3) = 273.74$	2.63
Error	29	$SCE(x_2, x_3) = 3016.59$	$CME(x_2, x_3) = 104.02$	
Total	31	$STC = 3564.08$	$f_{0.95,2,29} = 3.33$	

suma total de cuadrados es STC = 3 564.08 sin importar cuántas variables de predicción se incluyan en el modelo. Para la regresión de Y sobre x_2, $SCR(x_2) = 525.74$ es la reducción en la suma de los cuadrados de los errores cuando se añade el término $\beta_2 x_2$ al modelo $Y_i = \beta_0 + \varepsilon_i$. En otras palabras, si se ajusta el modelo $Y_i = \beta_0 + \varepsilon_i$, se supone que la única fuente de variación en Y_i es el error aleatorio; la recta de regresión estimada es simplemente $\hat{Y}_i = \bar{Y}$. Cuando se agrega el término $\beta_2 x_2$ al modelo, entonces parte de la variación total puede explicarse por la presencia de x_2. Esto es lo que precisamente representa $SCR(x_2) = 525.74$, $SCR(x_2)$ es la suma extra de cuadrados en la que disminuye SCE cuando se añade el término $\beta_2 x_2$ al modelo. Al emplear el mismo argumento para la regresión de Y sobre x_3, $SCR(x_3) = 353.70$ es la suma extra de cuadrados en los que disminuye el error cuando se añade el término $\beta_3 x_3$ al modelo $Y_i = \beta_0 + \varepsilon_i$. Para cualquier otro caso, si la reducción en la suma de los cuadrados de los errores es substancial, se rechaza la hipótesis nula de valor cero para el correspondiente coeficiente de regresión. Nótese que se rechaza x_2, H_0: $\beta_2 = 0$ para la regresión de Y sobre x_2 (valor $f = 5.19 > f_{0.95,1,30} = 4.17$), pero para la regresión de Y sobre x_3, H_0: $\beta_3 = 0$ no puede rechazarse.

Considérese la regresión de Y sobre x_2 y x_3. Lo que se desea determinar es la reducción en la suma de los cuadrados de los errores cuando se añade el término $\beta_3 x_3$ al modelo, el cual ya contiene el término constante β_0 y el término $\beta_2 x_2$, o la reducción en SCE cuando se introduce el término $\beta_2 x_2$ al modelo, el cual ya contiene a β_0 y $\beta_3 x_3$. Nótese que para el modelo *c* de la tabla 14.3 la suma de los cuadrados de los errores cuando se incluye en el modelo de regresión, tanto a x_2 como a x_3 es $SCE(x_2, x_3) = 3\ 016.59$. Pero cuando sólo se tiene a x_2 en el modelo, $SCE(x_2) = 3\ 038.34$. Por lo tanto, la diferencia entre $SCE(x_2)$ y $SCE(x_2, x_3)$ debe ser la suma de cuadrados extra debida a la inclusión del término $\beta_3 x_3$ en el modelo que ya contiene a los términos β_0 y $\beta_2 x_2$. Se denotará esta diferencia por $\text{SCR}(x_3 \mid x_2)$. De esta forma

$$\begin{aligned} \text{SCR}(x_3 \mid x_2) &= \text{SCE}(x_2) - \text{SCE}(x_2, x_3) \\ &= 3038.34 - 3016.59 \\ &= 21.75 \end{aligned} \tag{14.14}$$

es la reducción adicional en la suma de los cuadrados de los errores cuando se introduce x_3 en el modelo que ya contiene a x_2.

Dado que una reducción en la suma de los cuadrados de los errores significa un aumento correspondiente a la suma de los cuadrados de la regresión,

$$\begin{aligned} \text{SCR}(x_2, x_3) &= \text{SCR}(x_3 \mid x_2) + \text{SCR}(x_2) \\ &= 21.75 + 525.74 \\ &= 547.49. \end{aligned} \tag{14.15}$$

La suma de los cuadrados de la regresión, cuando figuran en el modelo, tanto x_2 como x_3, se separa en dos componentes, cada uno de éstos con un grado de libertad. $SCR(x_3 \mid x_2)$, el cual refleja la contribución de x_3 cuando ésta se añade al modelo $Y = \beta_0 + \beta_2 x_2 + \varepsilon$, y $SCR(x_2)$ la cual mide la contribución de x_2 cuando ésta se añade al modelo $Y = \beta_0 + \varepsilon$.

TABLA 14.4 Tabla ANOVA aumentada para la regresión de Y sobre x_2 y x_3

Fuente de variación	*gl*	*SC*	*CM*	*Valor F*
Regresión	2	SCR $(x_2, x_3) = 547.49$	CMR $(x_2, x_3) = 273.74$	2.63
x_2	1	SCR $(x_2) = 525.74$	CMR $(x_2) = 525.74$	5.05
$x_3 \mid x_2$	1	SCR$(x_3 \mid x_2) = 21.75$	CMR$(x_3 \mid x_2) = 21.75$	0.2
Error	29	SCE$(x_2, x_3) = 3016.59$	CME$(x_2, x_3) = 104.02$	
Total	31	STC = 3564.08	$f_{0.95,2,29} = 3.33$; $f_{0.95,1,29} = 4.18$	

Se puede demostrar que $SCR(x_3 \mid x_2)$ y $SCR(x_2)$ son variables aleatorias independientes chi-cuadrada, cada una con un grado de libertad; entonces puede hacerse una comparación entre el cuadrado medio correspondiente a $SCR(x_3 \mid x_2)$, o el de $SCR(x_2)$, y el cuadrado medio del error, $CME(x_2, x_3)$ por medio de la estadística F. Esta prueba se conoce como *prueba F parcial* sobre una variable de predicción. En realidad, la *prueba F parcial* determina si la contribución de un coeficiente de regresión es lo suficientemente grande para garantizar su inclusión en el modelo, dado que otros términos no toman en cuenta al coeficiente que ya se encuentra en el mismo. Por lo tanto, en cierto sentido, se intenta enjuiciar el efecto individual de la correspondiente variable de predicción sobre una respuesta dada. La tabla ANOVA aumentada para la regresión de Y sobre x_2 y x_3, la cual incluye las pruebas F parciales, se muestra en la tabla 14.4. Nótese que la inclusión del término $\beta_2 x_2$ en el modelo $Y = \beta_0 + \varepsilon$ tiene un efecto benéfico, mientras que la inclusión de $\beta_3 x_3$ en $Y = \beta_0 + \beta_2 X_2 + \varepsilon$ no.

A lo largo de toda la presentación anterior se supuso que el término $\beta_3 x_3$ era el último en sumarse al modelo que incluye a x_2 y x_3. Sin embargo, es posible realizar pruebas parciales F para cada coeficiente de regresión, como si la correspondiente variable de predicción fuese la última en haberse añadido al modelo. De esta forma, los efectos individuales de cada variable de predicción, en presencia de las otras, pueden comprobarse. Para el ejemplo, lo que se desea es determinar la contribución del término $\beta_2 x_2$ cuando el modelo ya contiene a β_0 y a $\beta_3 x_3$.

Al seguir el mismo procedimiento dado con anterioridad, la suma de los cuadrados de los errores cuando, tanto x_2 como x_3 se encuentran en el modelo, es $SCE(x_2, x_3) = 3\,016.59$. Pero cuando sólo se encuentra x_3 en el modelo, $SCE(x_3) = 3\,210.38$. De esta forma, la reducción en el valor de la suma de los cuadrados de los errores cuando se añade el término $\beta_2 x_2$ al modelo que ya contiene a β_0 y $\beta_3 x_3$ es

$$\begin{aligned} \text{SCR}(x_2 \mid x_3) &= \text{SCE}(x_3) - \text{SCE}(x_2, x_3) \qquad (14.16) \\ &= 3210.38 - 3016.59 \\ &= 193.79. \end{aligned}$$

Entonces la suma de los cuadrados de regresión, cuando x_2 y x_3 se encuentran en el

modelo, la suma de los dos componentes es

$$\begin{aligned} SCR(x_2, x_3) &= SCR(x_2 \mid x_3) + SCR(x_3) \\ &= 193.79 + 353.70 \\ &= 547.49, \end{aligned} \tag{14.17}$$

cada componente con un grado de libertad. Una consecuencia importante de todo lo anterior, es que tanto $SCR(x_2 \mid x_3) = 193.79$ y $SCR(x_3 \mid x_2) = 21.75$ son más pequeños que $SCR(x_2) = 525.74$ y $SCR(x_3) = 353.70$, respectivamente. El porqué de lo anterior constituye el tema de la siguiente sección.

Para determinar las pruebas *F* parciales para la regresión debida a x_3, o a x_2 dada x_3, ahora es posible tener otra versión de la tabla 14.4; ésta se muestra en la tabla 14.5. Nótese que una comparación entre los resultados de las tablas 14.4 y 14.5 muestra un desacuerdo con respecto al efecto de x_2 sobre la producción de gasolina. Mientras que la regresión lineal simple de *Y* sobre x_2 es estadísticamente significativa ($f = 5.19$), la regresión de *Y* sobre x_2 dada la presencia de x_3, no lo es ($f = 1.86$). Se dará más información con respecto a esta ocurrencia en la siguiente sección.

El principio de la suma de cuadrados extra se extiende de manera directa para aplicar la idea básica a cualquier número de variables de predicción. Por ejemplo, supóngase que se tienen tres variables de predicción x_1, x_2 y x_3. Se puede definir la reducción en la suma de los cuadrados de los errores, cuando una de éstas se añade al modelo que ya contiene a las otras dos, de la siguiente manera:

$$SCR(x_3 \mid x_1, x_2) = SCE(x_1, x_2) - SCE(x_1, x_2, x_3), \tag{14.18}$$

$$SCR(x_2 \mid x_1, x_3) = SCE(x_1, x_3) - SCE(x_1, x_2, x_3), \tag{14.19}$$

$$SCR(x_1 \mid x_2, x_3) = SCE(x_2, x_3) - SCE(x_1, x_2, x_3). \tag{14.20}$$

Para desarrollar expresiones similares a (14.15) o (14.17), de (14.14) se deduce que

$$SCR(x_2 \mid x_1) = SCE(x_1) - SCE(x_1, x_2),$$

TABLA 14.5 Tabla ANOVA aumentada para la regresión de *Y* sobre x_2 y x_3

Fuente de variación	*gl*	*SC*	*CM*	*Valor F*
Regresión	2	$SCR(x_2, x_3) = 547.49$	$CMR(x_2, x_3) = 273.74$	2.63
x_3	1	$SCR(x_3) = 353.70$	$CMR(x_3) = 353.70$	3.40
$x_2 \mid x_3$	1	$SCR(x_2 \mid x_3) = 193.79$	$CMR(x_2 \mid x_3) = 193.79$	1.86
Error	29	$SCE(x_2, x_3) = 3016.59$	$CME(x_2, x_3) = 104.02$	
Total	31	STC = 3564.08	$f_{0.95, 2, 29} = 3.33$; $f_{0.95, 1, 29}$	= 4.18

o

$$\text{SCE}(x_1, x_2) = \text{SCE}(x_1) - \text{SCR}(x_2 \mid x_1). \tag{14.21}$$

Ahora, cuando sólo se tiene a x_1 en el modelo, por definición

$$\text{SCE}(x_1) = \text{STC} - \text{SCR}(x_1);$$

pero cuando todas las tres variables se encuentran en el modelo,

$$\text{STC} = \text{SCR}(x_1, x_2, x_3) + \text{SCE}(x_1, x_2, x_3).$$

Entonces

$$\text{SCE}(x_1) = \text{SCR}(x_1, x_2, x_3) + \text{SCE}(x_1, x_2, x_3) - \text{SCR}(x_1),$$

y al sustituir $SCE(x_1)$ en (14.21), se obtiene

$$\text{SCE}(x_1, x_2) = \text{SCR}(x_1, x_2, x_3) + \text{SCE}(x_1, x_2, x_3) - \text{SCR}(x_1) - \text{SCR}(x_2 \mid x_1). \tag{14.22}$$

Al sustituir (14.22) por $SCE(x_1, x_2)$ en (14.18) se obtiene el resultado deseado

$$\text{SCR}(x_3 \mid x_1, x_2) = \text{SCR}(x_1, x_2, x_3) - \text{SCR}(x_1) - \text{SCR}(x_2 \mid x_1), \tag{14.23}$$

o

$$\text{SCR}(x_1, x_2, x_3) = \text{SCR}(x_1) + \text{SCR}(x_2 \mid x_1) + \text{SCR}(x_3 \mid x_1, x_2). \tag{14.24}$$

La suma de los cuadrados de regresión, cuando las tres variables se encuentran en el modelo, tiene tres componentes, cada uno con un grado de libertad. $SCR(x_1)$ mide la contribución (reducción en la suma de los cuadrados de los errores) de x_1 cuando se añade x_1 al modelo $Y = \beta_0 + \varepsilon$; $SCR(x_2 \mid x_1)$ representa la contribución de x_2 cuando ésta se introduce al modelo $Y = \beta_0 + \beta_1 x_1 + \varepsilon$; y $\text{SCR}(x_3 \mid x_1, x_2)$ es la contribucion de x_3 cuando ésta se agrega al modelo $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$.

Al emplear (14.19) o (14.20) y si se sigue el mismo procedimiento, pueden establecerse resultados similares a (14.24) de la siguiente manera:

$$\text{SCR}(x_1, x_2, x_3) = \text{SCR}(x_1) + \text{SCR}(x_3 \mid x_1) + \text{SCR}(x_2 \mid x_1, x_3), \tag{14.25}$$

$$\text{SCR}(x_1, x_2, x_3) = \text{SCR}(x_2) + \text{SCR}(x_3 \mid x_2) + \text{SCR}(x_1 \mid x_2, x_3). \tag{14.26}$$

Estos resultados permiten que se lleven a cabo pruebas F parciales sobre cada coeficiente de regresión como si la variable de predicción asociada con éste hubiese sido la última en incluirse en el modelo. En otras palabras, con las pruebas parciales F puede determinarse si el efecto individual de una variable de predicción en presencia de las demás es estadísticamente discernible. Debe notarse que al intercambiar el orden de entrada al modelo para las variables de predicción, entonces es posible identificar otras relaciones similares a (14.24)-(14.26). Por ejemplo,

$$\text{SCR}(x_1, x_2, x_3) = \text{SCR}(x_2) + \text{SCR}(x_1 \mid x_2) + \text{SCR}(x_3 \mid x_2, x_1)$$

es otra separación de $SCR(x_1, x_2, x_3)$. Conforme crece el número de variables de predicción, el número posible de separaciones se vuelve más grande.

Con base en lo anterior, puede explicarse ahora, para los datos de Prater dados en la sección anterior, la información que aparece en la parte media de la figura 14.1. El lector notará dos columnas identificadas como "*SC* tipo I" y "*SC* tipo IV". La de tipo I contiene las cuatro componentes de $SCR(x_1, x_2, x_3, x_4)$, tales que

$$\text{SCR}(x_1, x_2, x_3, x_4) = \text{SCR}(x_1) + \text{SCR}(x_2 \mid x_1) + \text{SCR}(x_3 \mid x_1, x_2) + \text{SCR}(x_4 \mid x_1, x_2, x_3).$$

Cada componente tiene un grado de libertad y representa la reducción en la suma de los cuadrados de los errores cuando se añade al modelo la variable indentificada. El orden de entrada de variables al modelo es el mismo para el cual fueron identificadas las variables de predicción por el usuario, así que

$$\text{SCR}(x_1) = 216.26,$$
$$\text{SCR}(x_2 \mid x_1) = 309.85,$$
$$\text{SCR}(x_3 \mid x_1, x_2) = 29.21,$$
$$\text{SCR}(x_4 \mid x_1, x_2, x_3) = 2873.95.$$

Las dos columnas que se encuentran inmediatamente a la derecha de la columna que corresponde a "*SC* tipo I", dan los valores de las pruebas F parciales y los valores correspondientes p para cada una de las cuatro componentes. A partir de esta información, es claro que el efecto individual de cada coeficiente de regresión en presencia de otros términos en el modelo es estadísticamente apreciable.

La *SC* tipo IV representa la reducción en la suma de los cuadrados de los errores debida a la edición, en el modelo, de la variable de predicción correspondiente, dado que las otras tres ya se encuentran en el mismo. Para el ejemplo, las componentes son

$$\text{SCR}(x_1 \mid x_2, x_3, x_4) = 25.82,$$
$$\text{SCR}(x_2 \mid x_1, x_3, x_4) = 11.20,$$
$$\text{SCR}(x_3 \mid x_1, x_2, x_4) = 130.68,$$
$$\text{SCR}(x_4 \mid x_1, x_2, x_3) = 2873.95.$$

Nótese que no existe ninguna razón teórica para que la suma de estas cuatro componentes sea igual a $SCR(x_1, x_2, x_3, x_4)$.

Con base en los valores de las pruebas F parciales para estas componentes, es clara la existencia de cierta discrepancia entre estos resultados y los que se tienen para *SC* tipo 1. Por ejemplo, la contribución de x_2 en presencia sólo de x_1, es estadísticamente discernible, pero no puede decirse lo mismo de la contribución de x_2 en presencia de x_1, x_3 y x_4.

14.4 El problema de la multicolinealidad

Es muy común obtener conclusiones equivocadas con un punto de vista casual para la aplicación de análisis de regresión, cuando no se tiene una completa apreciación de los problemas que pueden encontrarse. En la sección anterior se notaron varias de las discrepancias que pueden presentarse en la regresión lineal múltiple. Éstas proporcionan información valiosa para identificar problemas que necesitan una atención adicional. El enfoque para el análisis de regresión no debe ser simplemente maximizar el coeficiente de correlación múltiple sin tomar en cuenta la debida consideración de los coeficientes de regresión estimados y sus desviaciones estándar, o la de comprobar las suposiciones fundamentales del análisis de regresión.

Un problema frecuente en regresión lineal múltiple es el que algunas de la variables de predicción están correlacionadas. Si la correlación es pequeña, las consecuencias serán de índole menor. Sin embargo, si existe una correlación muy fuerte entre dos o más variables de predicción, los resultados de la regresión serán ambiguos, especialmente con respecto a los valores de los coeficientes de regresión estimados. Un coeficiente de correlación muy alto entre dos o más variables de predicción constituye lo que se conoce como *multicolinealidad.* Este problema muchas veces es difícil de detectar ya que surge como consecuencia de datos deficientes. Éste es el precio que se paga cuando no es posible diseñar los experimentos en forma estadística y recabar los datos en arreglos balanceados, tal como se analizó en el capítulo 12.

Recuérdese que la ecuación de predicción, a pesar de que no es precisa en un sentido físico, debe ser un medio, empírico, viable para predecir la respuesta promedio dada una condición de las variables de predicción. La multicolinealidad no impide tener un buen ajuste ni evita que la respuesta sea, en forma adecuada, predicha dentro del intervalo de las observaciones; lo que sucede es que ésta afecta en forma severa las estimaciones de mínimos cuadrados, ya que bajo los efectos de la multicolinealidad éstas tienden a ser menos precisas para los efectos individuales de las variables de predicción, es decir, cuando dos o más variables de predicción son colineales los coeficientes de regresión estimados no miden los efectos individuales sobre la respuesta, sino que reflejan un efecto parcial sobre la misma, sujeto a todo lo que pase con las demás variables de predicción en la ecuación de regresión.

Para apreciar la naturaleza de la multicolinealidad, primero se estudiará una situación en la que ésta no existe. Considérese un modelo de regresión con dos variables de predicción. Si el coeficiente de correlación simple entre las dos variables es cero, entonces se dice que las variables son *ortogonales.** Al tener variables de predicción ortogonales el efecto que una de éstas tiene sobre la respuesta dada se mide en forma totalmente independiente del efecto individual que la otra variable tiene sobre la misma respuesta. Si una o ambas variables de predicción se encuentran en la ecuación de regresión, las estimaciones de mínimos cuadrados no cambiarán su valor

* Una de las principales razones para diseñar experimentos en forma estadística es la de adquirir factores o variables que sean ortogonales. Para muchos de los experimentos que emplean el análisis de varianza, los factores son ortogonales.

TABLA 14.6 Datos de la muestra para el ejemplo 14.2

Y(°F)	x_1(°F)	x_2(%)
66	70	20
72	75	20
77	80	20
67	70	30
73	75	30
78	80	30
68	70	40
74	75	40
79	80	40

Fuente: Servicio Climatológico Nacional.

Ejemplo 14.2 Para ilustrar los efectos ortogonales se examinarán los datos (limitados) que aparecen en la tabla 14.6 que consisten en la temperatura aparente Y (qué tan caliente se siente) como una función de la temperatura del aire x_1 y de la humedad relativa x_2.

El lector no tendrá ningún problema pra verificar que el coeficiente de correlación entre x_1 y x_2 tiene un valor de cero. Se procederá a ajustar los modelos $Y = \beta_0 + \beta_1 x_1 + \varepsilon$, $Y = \beta_0 + \beta_2 x_2 + \varepsilon$, y $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$. La información relevante se encuentra en la tabla 14.7.

Nótese que los coeficientes de regresión estimados para x_1 y x_2 son 1.10 y 0.10, respectivamente, sin importar que una o ambas variables de predicción se encuentren en la ecuación de regresión. De esta forma, por cada grado que aumenta la temperatura del aire, la temperatura aparente aumenta en 1.10 grados, y por cada incremento en porcentaje de la humedad relativa, la temperatura aparente aumenta 0.10 grados.* Además, nótese que

$$\text{SCR}(x_2 \mid x_1) = \text{SCR}(x_2),$$

$$\text{SCR}(x_1, x_2) = \text{SCR}(x_1) + \text{SCR}(x_2).$$

Los resultados anteriores son los que se esperan cuando las variables de predicción son ortogonales y no existe multicolinealidad.

Si se consideran de nuevo los datos de Prater y las regresiones que incluyen a x_2 o x_3, dadas en la tabla 14.3, se mostrará que existe una razón para sospechar la existencia de multicolinealidad entre x_2 y x_3. Primero, nótese que el coeficiente de regresión estimado para x_2 es 1.57 cuando sólo se encuentra presente en la ecuación de regresión x_2, pero su valor es de 2.26 cuando se añade x_3. De manera similar, el coeficiente de x_3 es -0.09 para el modelo de línea recta, pero éste cambia tanto en signo como en valor para ser igual a 0.05 cuando también se incluye a x_2 en la ecuación de regresión. Segundo, es claro que la reducción en el valor de la suma de los cuadrados

* El lector no debe generalizar de estos resultados por lo limitado de los datos.

TABLA 14.7 Ecuaciones de regresión estimadas y tablas ANOVA para la temperatura aparente, tomando a x_1 y/o x_2.

$$\hat{y} = -9.83 + 1.10x_1$$

Fuente de variación	*gl*	*SC*	*CM*	*Valor F*
Regresión	1	$SCR(x_1) = 181.5$	$CMR(x_1) = 181.5$	195.46
Error	7	$SCE(x_1) = 6.5$	$CME(x_1) = 0.9286$	
Total	8	$STC = 188.0$	$f_{0.95,1,7} = 5.59$	

$$\hat{y} = 69.67 + 0.10x_2$$

Fuente de variación	*gl*	*SC*	*CM*	*Valor F*
Regresión	1	$SCR(x_2) = 6.0$	$CMR(x_2) = 6.0$	0.23
Error	7	$SCE(x_2) = 182.0$	$CME(x_2) = 26.0$	
Total	8	$STC = 188.0$	$f_{0.95,1,7} = 5.59$	

$$\hat{y} = 12.83 + 1.10x_1 + 0.10x_2$$

Fuente de variación	*gl*	*SC*	*CM*	*Valor F*
Regresión	2	$SCR(x_1, x_2) = 187.5$	$CMR(x_1, x_2) = 93.75$	1125.0
x_1	1	$SCR(x_1) = 181.5$	$CMR(x_1) = 181.5$	2178.0
$x_2 \mid x_1$	1	$SCR(x_2 \mid x_1) = 6.0$	$CMR(x_2 \mid x_1) = 6.0$	72.0
Error	6	$SSE(x_1, x_2) = 0.5$	$CME(x_1, x_2) = 0.0833$	
Total	8	$STC = 188.0$	$f_{0.95,2,6} = 5.14$, $f_{0.95,1,6} = 5.99$	

de los errores debida a x_3 cuando x_2 se encuentra en el modelo, $SCR(x_3 \mid x_2) = 21.75$ es mucho menor que cuando sólo se encuentra x_3 en el modelo, $SCR(x_3) = 353.70$. La fuerte correlación que en forma aparente existe entre x_2 y x_3 ha disminuido de manera drástica el efecto individual que sobre la respuesta tiene x_3 en presencia de x_2. Puede hacerse el mismo comentario con respecto al efecto de x_2, ya que éste es estadísticamente apreciable en ausencia de x_3 ($SCR(x_2) = 525.74$, $f = 5.19$), pero se encuentra sustancialmente disminuido cuando x_3 se encuentra presente ($SCR(x_2 \mid x_3) = 193.79$).

Para mostrar que existe una fuerte correlación entre x_2 y x_3, se determinará la matriz de correlación para las cuatro variables de predicción de los datos de Prater. Esta matriz contiene todos los pares posibles de coeficientes de correlación y puede

determinarse para un conjunto dado de variables en forma muy fácil mediante el empleo de un paquete para computadora.* La matriz de correlación para x_1, x_2, x_3, y x_4 es la siguiente:

$$\begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{array} \begin{array}{c} \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} \\ \begin{bmatrix} 1.00 & 0.62 & -0.70 & -0.32 \\ 0.62 & 1.00 & -0.91 & -0.30 \\ -0.70 & -0.91 & 1.00 & 0.41 \\ -0.32 & -0.30 & 0.41 & 1.00 \end{bmatrix} \end{array}$$

Nótese que el valor de cada uno de los elementos que se encuentran en la diagonal es uno, ya que cada variable se encuentra correlacionada de manera perfecta consigo misma. Los elementos que se encuentran fuera de la diagonal son los valores de los coeficientes de correlación simple. Por ejemplo, $r_{12} = 0.62$ es el coficiente de correlación entre x_{10} y x_2; por lo tanto, el valor $r_{23} = -0.91$ al encontrarse muy cercano a -1 sugiere una fuerte asociación lineal entre x_2 y x_3. Este resultado es predecible si se inspeccionan en forma visual los datos dados en el ejemplo 14.1. Nótese que conforme aumenta la presión de vapor del petróleo crudo x_2, el punto x_3 *ASTM* 10% disminuye y viceversa. Estos resultados proporcionan la causa para sospechar la presencia de multicolinealidad en este ejemplo.

¿Qué es lo que se puede hacer cuando se descubre la presencia de multicolinealidad? Una alternativa es la de añadir puntos de observación para las variables colineales, los cuales tiendan a disminuir la severidad de la correlación. Pero puede ocurrir que estos puntos de observación no se encuentren disponibles fácilmente. Por ejemplo, para los datos de la gasolina podrían no existir los tipos de petróleo crudo que pueden disminuir la fuerte linealidad que existe entre x_2 y x_3. Una segunda alternativa es la de omitir una o más de las variables que son colineales, lo que reduce la variabilidad de los coeficientes de regresión de las restantes variables. Se han desarrollado enfoques más sofisticados para resolver los problemas que plantea la multicolinealidad, incluyendo la regresión por componentes principales y la regresión ridge. Estos temas se encuentran más allá del objetivo de este libro; se invita al lector a que consulte las referencias [1] y [3].

Para ilustrar la segunda alternativa y resolver el problema de la multicolinealidad, se examinarán las regresiones para las cuales se omiten x_2 o x_3. Como comparación, también se considerará la regresión de la producción de gasolina con respecto sólo al punto (x_3) *ASTM* 10% y al punto final (x_4). Sin proporcionar argumentación adicional se piensa que estas tres regresiones son las candidatas para la "mejor" ecuación de regresión lineal para los datos de Prater. La información más importante se encuentra en la tabla 14.8.

Al comparar parece que la regresión de Y sólo sobre x_3 y x_4 es la mejor con respecto a las proporcionadas por los otros dos modelos. Para el modelo *b*, la desviación estándar del estimador por mínimos cuadrados para el término constante es muy gran-

* Para *SAS* puede ser apropiado utilizar *PROC CORR*.

TABLA 14.8 Candidatos para la mejor ecuación de regresión para los datos de Prater

a) Regresión de Y sobre x_1, x_2, x_4

Variable	*Coeficiente de regresión estimado*	*Desviación estándar estimada*	*Valor t*
Constante	−53.899	5.8135	−9.27
x_1	0.422	0.1273	3.32
x_2	2.154	0.2716	7.93
x_4	0.144	0.0084	17.10

$R^2 = 0.9255$ $t_{0.975,28} = 2.048$

ANOVA

Fuente	*gl*	*SC*	*CM*	*Valor F*
Regresión	3	3298.60	1099.53	115.97
x_1	1	216.26	216.26	22.81
$x_2 \mid x_1$	1	309.85	309.85	32.68
$x_4 \mid x_1, x_2$	1	2772.49	2772.49	292.41
Error	28	265.48	9.48	
Total	31	3564.08	$f_{0.95,3,28} = 2.95$; $f_{0.95,1,28} \doteq 4.20$	

b) Regresión de Y sobre x_1, x_3, x_4

Variable	*Coeficiente de regresión estimado*	*Desviación estándar estimada*	*Valor t*
Constante	4.032	7.2233	0.56
x_1	0.222	0.1021	2.17
x_3	−0.187	0.0159	−11.72
x_4	0.157	0.0065	24.22

$R^2 = 0.959$ $t_{0.975,28} = 2.048$

ANOVA

Fuente	*gl*	*SC*	*CM*	*Valor F*
Regresión	3	3418.08	1139.38	218.51
x_1	1	216.26	216.26	41.47
$x_3 \mid x_1$	1	142.08	142.08	27.25
$x_4 \mid x_1, x_3$	1	3059.74	3059.74	586.79
Error	28	146.00	5.21	
Total	31	3564.08	$f_{0.95,3,28} = 2.95$; $f_{0.95,1,28} = 4.20$	

(continúa)

TABLA 14.8 *(continuación)*

c) Regresión de Y sobre x_3, x_4

Variable	*Coeficiente de regresión estimado*	*Desviación estándar estimada*	*Valor t*
Constante	18.468	3.0090	6.14
x_3	−0.209	0.0127	−16.43
x_4	0.156	0.0069	22.73

$R^2 = 0.9521$ $t_{0.975,\,29} = 2.045$

ANOVA

Fuente	*gl*	*SC*	*CM*	*Valor F*
Regresión	2	3393.47	1696.73	228.41
x_3	1	353.70	353.70	60.12
$x_4 \mid x_3$	1	3039.77	3039.77	516.69
Error	29	170.61	5.88	
Total	31	3564.08	$f_{0.95,\,2,\,29} = 3.33$; $f_{0.95,\,1,\,29} = 4.18$	

de, y la desviación estándar del coeficiente x_1 es casi igual a la mitad del valor de éste. Para el modelo *a*, $R^2 = 0.9255$, mientras que para el modelo *c*, $R^2 = 0.9521$, el cual es un valor mucho más cercano al valor de R^2 cuando todas las variables de predicción figuran en la ecuación de regresión. Además, las desviaciones estándar de los coeficientes de regresión estimados son, en forma relativa, mejores para el modelo *c* que para el *b*. Para finalizar, los factores físicos claves como la consistencia lógica de los coeficientes de regresión estimados, son los que por lo general definen la elección final.

14.5 Determinación del mejor conjunto de variables de predicción

Un problema muy importante en el análisis de regresión es determinar cuáles de las variables de predicción en la lista inicial deberán incluirse en el modelo de regresión. En casi todas las ocasiones, un investigador decidirá, de una lista inicial de variables de predicción, a aquéllas que tienen la mayor probabilidad de contener los factores más importantes para la respuesta dada. Por lo tanto, es necesario tener una manera de determinar, de la lista inicial de variables de predicción, a aquéllas que parecen ser las mejores para describir el cambio en la respuesta promedio, y de esta forma proporcionarán una ecuación de predicción representativa de las condiciones bajo las cuales se recabaron los datos. La palabra "mejores" no debe interpretarse como poseedora de la connotación teórica de óptimo; ésta debe considerarse como representativa de los medios por los cuales se aíslan las características más sobresalientes, de tal manera que puede llevarse a cabo un análisis significativo.

Sea k el número inicial de potenciales variables de predicción; el número de términos en el modelo lineal completo, incluyendo al término constante, es $m = k + 1$. Un procedimiento que es muy recomendable para determinar el mejor conjunto de variables de predicción por incluir en la ecuación de regresión es calcular y comparar todas las posibles 2^k ecuaciones de regresión. Con este proceso se tendrá una ecuación, la cual no contiene ninguna variable de predicción ($\hat{Y} = \bar{Y}$), k ecuaciones cada una con una variable de predicción, $k(k - 1)/2$ ecuaciones con dos variables de predicción y así sucesivamente. El procedimiento proporciona al investigador la oportunidad de evaluar y comparar todas las ecuaciones de regresión y, con base en la investigación de todas las discrepancias aparentes, debe surgir la mejor ecuación. Dado que hoy en día la capacidad de cómputo es muy extensa, la determinación de todas las posibles ecuaciones de regresión es el mejor método, aun si k tiene un valor tan grande como 9 o 10.

Cuando k es grande, puede no ser práctico determinar y evaluar todas las posibles ecuaciones de regresión. Para estos casos, se han desarrollado *técnicas para la selección de las variables* que pueden proporcionar al usuario información muy útil, sin tener que evaluar todas las posibles ecuaciones de regresión. Sin embargo, estas técnicas tienen algunos inconvenientes y no deben considerarse como iguales con respecto a la evaluación de todas las posibles regresiones. Mientras que los procedimientos para la selección de variables dan resultados confiables, cuando la muticolinealidad no es problema, éstos producirán resultados contradictorios para datos colineales. De esta forma, si se sospecha la presencia de multicolinealidad, no deberán emplearse métodos para la selección de variables. La técnica más usual de selección de variables emplea un procedimiento de *regresión por pasos* para obtener la mejor ecuación de regresión. Existen dos versiones principales de esta técnica: la selección hacia adelante y la eliminación hacia atrás.

El procedimiento de selección hacia adelante comienza con una ecuación que no contiene variables de predicción. La primera variable incluida en la ecuación es aquella que produce la mayor reducción en el valor de la suma de los cuadrados de los errores; ésta es la variable de predicción con el coeficiente de correlación simple más alto para la respuesta dada. Con base en una prueba de hipótesis, si el coeficiente de regresión es significativamente diferente de cero, la variable permanece en la ecuación y se comienza la búsqueda de una segunda variable. La segunda variable por incluir en la ecuación es aquella que produce la mayor reducción en la suma de los cuadrados de los errores, dada la presencia de la primera variable.* Ésta es la variable que posee el coeficiente de correlación más alto con la respuesta, después de que ésta se ha ajustado para tomar en cuenta el efecto de la primera variable. Si la significancia estadística es discernible para el coeficiente de regresión de la segunda variable, ésta se mantiene en la ecuación y se comienza la búsqueda de una tercera variable de predicción. El proceso se continúa de esta forma hasta que la significancia estadística no sea discernible para el coeficiente de la última variable que ha entrado a la ecuación.

El procedimiento de eliminación hacia atrás comienza con la ecuación de regresión que contiene a todas las variables de predicción. Entonces se eliminan, una a la

* En este momento pueden surgir dificultades cuando los datos son colineales.

vez, las variables menos importantes con base en su contribución a la reducción en el valor de la suma de los cuadrados de los errores. Por ejemplo, la primera variable por omitir será aquella cuyo efecto sobre la reducción en el valor de la suma de los cuadrados de los errores, dada la presencia de las demás variables, sea el más pequeño. El procedimiento concluye cuando los coeficientes de todas las variables que aún permanecen en la ecuación tienen una significancia estadísticamente discernible.

El procedimiento de selección hacia adelante se ha modificado de tal manera que se considere la posibilidad de eliminar una variable en cada etapa. Esta modificación da origen a lo que en forma usual se conoce en los paquetes de computación como procedimiento de regresión por pasos *(stepwise)*. Con este método puede eliminarse, en una etapa posterior, una variable de predicción cuya inclusión se llevó a cabo en una etapa anterior. De nuevo, el proceso de decisión se basa en la reducción en el valor de la suma de los cuadrados de los errores y de las pruebas *F* parciales y depende de la combinación particular de las variables que se tienen en la ecuación de regresión.

Con el desarrollo de paquetes para computadora cada vez más elaborados se tienen disponibles otras técnicas, pero la característica común sigue siendo el valor de la suma de los cuadrados de los errores cuando una variable entra a (o es removida de) la regresión, dada la presencia de las demás variables de predicción. Para datos "con buen comportamiento", los procedimientos de regresión por pasos y de eliminación hacia atrás en general proporcionan los mismos resultados. Si existe alguna diferencia entre éstos, este hecho muchas veces constituye una buena indicación para considerar el problema con mayor cuidado, así como la realización de análisis adicionales.

Para evaluar y comparar las ecuaciones de regresión, de manera especial dentro del contexto de todas las posibles regresiones, es necesario tener criterios efectivos. Dos de los criterios más útiles son el del cuadrado medio del error (*CME*) y el criterio C_p. Con el propósito de tener un panorama más completo, también se estudiará el coeficiente de correlación múltiple R^2.

1. *El criterio del cuadrado medio del error.* Recuérdese que el cuadrado medio del error es igual a la varianza residual. Dado que *CME* es la suma de los cuadrados de los residuos dividida entre el número de grados de libertad de *SCE*, *CME* toma en cuenta el número de parámetros en el modelo a través del número de grados de libertad. Mientras que la suma de los cuadrados de los errores no puede aumentar si se permiten más variables en el modelo, no ocurre lo mismo con el cuadrado medio del error si la reducción en el valor de *SCE* es tan pequeña que no pueda compensar la pérdida del número de grados de libertad adicionales. Por ejemplo, recuérdese la tabla 14.3 y en particular los modelos *a* y *c*. Nótese que $SCE(x_2) = 3038.34$ es mayor que $SCE(x_2, x_3) = 3016.59$, pero $CME(x_2) = 101.28$ es menor que $(x_2, x_3) = 104.02$. Con el criterio *CME* puede determinarse el conjunto de variables de predicción que minimice a *CME* o casi lo haga en el momento para el que la introducción de más variables de predicción en la ecuación de regresión ya no se encuentre garantizada.

2. *El criterio* C_p. Recuérdese que la varianza residual S^2 es un estimador no sesgado de la varianza del error σ^2 sólo cuando se ha escogido la forma correcta para el

modelo de regresión. De otra forma, puede demostrarse que

$$E(S^2) = \sigma^2 + \frac{\sum_{i=1}^{n} A_i^2}{(n - p)}, \tag{14.27}$$

donde p es el número de términos que aparecen en el modelo, incluido el término constante y

$$A_i = E(Y_i) - E(\hat{Y}_i)$$

es el sesgo.

Supóngase que la ecuación de regresión, la cual contiene k variables de predicción, se ha escogido en forma cuidadosa, de tal manera que $CME \equiv S^2$ es un estimador no sesgado de σ^2. Pero para cualquier ecuación de regresión que sólo contenga a un subconjunto de las k variables de predicción, es posible que $A_i \neq 0$, y las predicciones de la respuesta con base en la ecuación de regresión estimada pueden encontrarse sesgadas. Para evaluar la efectividad de esta ecuación de regresión, como un medio para formular predicciones, debe considerarse el cuadrado medio del error de un valor predicho, más que la varianza de éste. El cuadrado medio del error total estandarizado que se define como

$$\begin{aligned} \Gamma_p &= \frac{1}{\sigma^2} \sum_{i=1}^{n} \mathrm{CME}(\hat{Y}_i) \\ &= \frac{1}{\sigma^2} \left[\sum_{i=1}^{n} A_i^2 + \sum_{i=1}^{n} Var(\hat{Y}_i) \right], \end{aligned} \tag{14.28}$$

se ha propuesto como un criterio apropiado de la bondad del ajuste para una ecuación de regresión estimada la cual contiene p términos. La cantidad Γ_p considera tanto a la componente del sesgo en $\hat{Y}_i$, ya que algunas de las variables de predicción no se encuentran incluidas, así como a la varianza en $\hat{Y}_i$ para todas las n observaciones de la respuesta. A continuación se obtendrá un estimador para Γ_p.

Puede demostrarse que

$$\sum_{i=1}^{n} Var(\hat{Y}_i) = p\sigma^2,$$

lo cual implica que la varianza total de la predicción aumenta conforme el número de términos en la ecuación de regresión también aumenta. Al sustituir en (14.28), se tiene

$$\Gamma_p = \frac{1}{\sigma^2} \sum_{i=1}^{n} A_i^2 + p. \tag{14.29}$$

Dado que para una ecuación de regresión que contiene p términos

$$\mathrm{SCE}_p = (n - p)S_p^2,$$

se tiene

$$\begin{aligned} E(\mathrm{SCE}_p) &= (n-p)E(S_p^2) \\ &= (n-p)\left[\sigma^2 + \frac{\sum A_i^2}{(n-p)}\right] \\ &= (n-p)\sigma^2 + \sum A_i^2, \end{aligned}$$

o

$$\sum A_i^2 = E(\mathrm{SCE}_p) - (n-p)\sigma^2.$$

Al sustituir en (14.29), se obtiene

$$\begin{aligned} \Gamma_p &= \frac{E(\mathrm{SCE}_p) - (n-p)\sigma^2}{\sigma^2} + p \\ &= \frac{E(\mathrm{SCE}_p)}{\sigma^2} - (n-p) + p \\ &= \frac{E(\mathrm{SCE}_p)}{\sigma^2} - (n-2p). \end{aligned}$$

Dado que SCE_p es un estimador de $E(\mathrm{SCE}_p)$ y S_k^2 lo es a su vez de σ^2, un estimador de Γ_p es la estadística

$$C_p = \frac{\mathrm{CSE}_p}{S_k^2} - (n-2p). \tag{14.30}$$

Nótese que SCE_p es la suma de los cuadrados de los errores para la ecuación de regresión, la cual contiene p términos y que $S_k^2 = \mathrm{CME}(x_1, x_2, \ldots, x_k)$ es el estimador de σ^2 basado en todas las k variables de predicción.

Los valores deseables para C_p para la bondad del ajuste de una ecuación de re gresión que contiene p términos son aquellos que se encuentran muy cercanos a p. Lo anterior surge del hecho de que si el sesgo de una ecuación de regresión de p términos es despreciable $\Sigma A_i^2 \approx 0$ y $E(\mathrm{SCE}_p) = (n-p)\sigma^2$. Bajo esta condición, el valor esperado de la estadística C_p es

$$\begin{aligned} E(C_p \mid A_i = 0) &= \frac{(n-p)\sigma^2}{\sigma^2} - (n-2p) \\ &= p. \end{aligned}$$

De esta forma, cuando se obtienen todas las posibles regresiones, se calcula un valor de C_p para cada caso. Las regresiones que tienen valores de C_p cercanos a p se consideran como deseables.

Puede ser benéfico aceptar un pequeño sesgo en la predicción, mediante la eliminación de algunas variables de predicción, aun si sus coeficientes de regresión son es-

tadísticamente significativos, con excepción de los que tienen un valor igual a cero. Lo anterior es especialmente cierto si los coeficientes de regresión del nuevo modelo se estiman con varianzas pequeñas; además, dado que la varianza total de la predicción aumenta conforme se añaden más variables al modelo de regresión, puede ser ventajoso eliminar algunas variables con el propósito de disminuir el error promedio de la predicción.

Además de considerar a *CME* y a C_p, también se debe considerar el coeficiente de correlación múltiple R^2 para evaluar las ecuaciones de regresión. Dado que R^2 varía en forma inversa a como lo hace la suma de los cuadrados de los errores, R^2 aumentará conforme se añadan más variables al modelo de regresión y R^2 alcanzará su valor máximo cuando todas las variables de predicción se encuentren en la ecuación de regresión. Por lo tanto, la razón para emplear a R^2 como un criterio, no es la de encontrar el conjunto de variables que maximiza R^2, sino más bien determinar el punto más allá del cual sumar más variables no es deseable, ya que el incremento que se tiene en R^2 es mínimo.

Para proporcionar una ilustración de todas las posibles regresiones y sus comparaciones, tomando en cuenta los criterios anteriores, de nuevo considérense los datos de Prater. La tabla 14.9 contiene las estimaciones por mínimos cuadrados para los coeficientes de cada regresión (distintas de la trivial $\hat{y}_i = \bar{y} = 19.66$), y la tabla 14.10 identifica los correspondientes valores de SCE, CME, C_p y R^2.

El cuadrado medio del error cuando las cuatro variables de predicción se encuentran en el modelo de regresión es $\text{CME}(x_1, x_2, x_3, x_4) = 4.99$. De esta forma, por ejemplo, para obtener el valor de C_p para la regresión de Y sobre x_1, x_3, y x_4, se tiene que $SCE(x_1, x_3, x_4) = 146.00$, $p = 4$, $n = 32$ y

$$C_p = \frac{146}{4.99} - (32 - 8) = 5.26.$$

TABLA 14.9 Todas las regresiones posibles para los datos de Prater

Variables de predicción en el modelo	b_0	b_1	b_2	b_3	b_4
x_1	1.264	0.469			
x_2	13.087		1.572		
x_3	41.389			−0.090	
x_4	−16.662				0.019
x_1, x_2	12.256	0.025	1.539		
x_1, x_3	35.174	0.096		−0.080	
x_1, x_4	−64.951	1.009			0.136
x_2, x_3	− 2.524		2.257	0.053	
x_2, x_4	−37.808		2.677		0.139
x_3, x_4	18.468			−0.209	0.156
x_1, x_2, x_3	−11.013	0.125	2.278	0.067	
x_1, x_2, x_4	−53.899	0.422	2.154		0.144
x_1, x_3, x_4	4.032	0.222		−0.187	0.157
x_2, x_3, x_4	8.562		0.523	−0.175	0.154
x_1, x_2, x_3, x_4	− 6.821	0.227	0.554	−0.150	0.155

TABLA 14.10 Criterios de bondad de ajuste para todas las posibles regresiones para los datos de Prater

Variables de predicción	R^2	SCE	CME	C_p
x_1	0.0607	3347.82	111.59	642.91
x_2	0.1475	3038.34	101.28	580.89
x_3	0.0992	3210.38	107.01	615.36
x_4	0.5063	1759.69	58.66	324.64
x_1, x_2	0.1476	3037.97	104.76	582.81
x_1, x_3	0.1005	3205.74	110.54	616.43
x_1, x_4	0.7582	861.95	29.72	146.74
x_2, x_3	0.1536	3016.59	104.02	578.53
x_2, x_4	0.8962	369.87	12.75	48.12
x_3, x_4	0.9521	170.61	5.88	8.19
x_1, x_2, x_3	0.1558	3008.76	107.46	578.96
x_1, x_2, x_4	0.9255	265.48	9.48	29.20
x_1, x_3, x_4	0.9590	146.00	5.21	5.26
x_2, x_3, x_4	0 9549	160.62	5.74	8.19
x_1, x_2, x_3, x_4	0.9622	134.80	4.99	5.00

Al tomar en cuenta, tanto a *CME* como C_p, la mejor ecuación de predicción para la producción de gasolina debe seleccionarse de las regiones que incluyen (x_3, x_4), (x_1, x_3, x_4), (x_2, x_3, x_4), y (x_1, x_2, x_3, x_4). Esta última no es en particular atractiva, ya que las estimaciones de los coeficientes de regresión para el término constante y para x_2 tienen desviaciones estándar muy grandes. A pesar de que la ecuación de regresión que contenga x_2, x_3, y x_4 tiene valores de CME y C_p muy cercanos a los óptimos, ésta carece de una precisión satisfactoria para la estimación del coeficiente de x_2, dado que $b_2 = 0.523$, con $s(B_2) = 0.396$. Puede decirse lo mismo de la regresión que comprenda a x_1, x_3, y x_4 para las estimaciones de β_0 y del coeficiente de x_1 (véase el modelo *b* en la tabla 14.8). De acuerdo con lo anterior, se acepta un pequeño sesgo en la predicción y se concluye que la ecuación de regresión que contiene a x_3 y a x_4 es la mejor para predecir la producción de gasolina en el intervalo de valores de las observaciones.

A continuación se dan las etapas por seguir en un procedimiento de regresión paso a paso:

1. El procedimiento comienza mediante la obtención de *k* ecuaciones de regresión lineal simples.
   La estadística *F*

$$F = \text{CMR}(x_i)/\text{CME}(x_i)$$

se calcula para cada $i = 1, 2, \ldots, k$ variables. Si el mayor valor *F* excede un nivel de significancia estadística, previamente determinado, la variable correspondiente es la primera que se incluye en la regresión. De otro modo, la mejor ecuación es $\hat{Y} = \bar{Y}$. Este proceso es idéntico al que se sigue para determinar la variable de predicción que tiene la mayor correlación con la respuesta.

2. Supóngase que la variable x_3 entra a la ecuación de regresión en el paso 1. En este momento, el procedimiento de regresión paso a paso calcula todas las ecuaciones que contienen dos variables, incluyendo a x_3. Para cada caso, el valor de la estadística F parcial

$$F = \text{CMR}(x_i \mid x_3)/\text{CME}(x_i, x_3)$$

se calcula para determinar si puede rechazarse H_0: $\beta_i = 0$ en presencia de x_3. Si el mayor valor de F es suficiente para la significancia estadística, la segunda variable correspondiente se añade a la ecuación.
3. Supóngase que se añade x_1 a la ecuación en el paso 2. El procedimiento continúa mediante un examen para determinar si alguna de las otras variables que ya se encuentran en la ecuación debe eliminarse ahora; en este caso, ésta podría ser x_3. Se calcula el valor de la estadística F parcial

$$F = \text{CMR}(x_3 \mid x_1)/\text{CME}(x_1, x_3)$$

y se compara con un nivel predeterminado de significancia. Si el efecto de x_3 dado x_1 no es estadísticamente discernible, se elimina a x_3 de la ecuación; de otro modo se retiene. Para etapas posteriores existirá un cierto número de las pruebas F parciales para todas las variables que se añadieron en etapas anteriores. La variable que puede eliminarse es aquella para la que el valor de F es el más pequeño.
4. Supóngase que se retiene a x_3; en este momento la ecuación de regresión incluye a x_1 y a x_3. El proceso se continúa mediante un examen para determinar cuál de las variables restantes es candidata para incluirse en el modelo. Entonces, se examina si alguna de las variables que ya se encuentran incluidas debe eliminarse ahora. El proceso termina cuando ninguna de las demás variables de predicción puede añadirse o eliminarse del modelo de regresión.

Se deja como un ejercicio para el lector emplear los datos de producción de gasolina con todas las opciones de selección posibles de variables y se compararán los resultados.

14.6 Análisis de residuos o residuales

En la sección anterior se examinaron algunas formas para determinar la "mejor" ecuación de regresión, bajo las circunstancias impuestas por el conjunto de datos. Una manera muy efectiva de detectar las posibles deficiencias de un modelo radica en llevar a cabo un análisis de residuos. Ningún otro aspecto es tan importante en el análisis de regresión como el análisis de los residuos. El conocido economista Paul A. Samuelson comentaba: "al científico que hace predicciones le recomiendo que siempre estudie sus residuales".

Como se hizo notar en el capítulo 12, el análisis de los residuos puede descubrir las violaciones de las suposiciones o las deficiencias del modelo. Se examinarán tres deficiencias muy comunes: la ecuación de regresión puede no ser lineal en las variables de predicción; la varianza del error σ^2 puede no ser constante y una o más de las variables de predicción que ejercen una influencia importante pueden no estar in-

cluidas en el modelo. También se considerará el problema de las observaciones *discrepantes o aberrantes,* que son aquellas cuyos valores se encuentran alejados del comportamiento general del resto de los datos.

Recuérdese que el *i*-ésimo residuo e_i es la diferencia numérica que existe entre el valor observado y_i y el correspondiente valor estimado $\hat{y}_i$, para toda $i = 1, 2, \ldots, n$. El residuo e_i se considera como una estimación del verdadero error no observable ε_i. El error cuadrático medio es la varianza de los residuos, la que a su vez es una estimación de σ^2.

En esencia, el análisis de residuos significa realizar un análisis de sus gráficas de los residuos. Si se ha definido la ecuación de regresión en forma correcta y no existe ninguna deficiencia, entonces una gráfica de los residuos contra cualesquiera de los valores estimados $\hat{y}_i$ a los correspondientes valores de cada variable de predicción en la ecuación no mostrará ningún patrón, es decir, no existirá ninguna relación entre los residuos y los valores ajustados o entre los residuos y ios valores de las variables de predicción. Si existe alguna relación, ésta sugerirá el hecho de que hay una deficiencia en la ecuación de regresión. Para detectar las áreas de problemas a través del análisis de los residuos, es preferible, de nuevo, emplear los residuos estandarizados. Dado que la media de los residuos es igual a cero,

$$e_{i_s} = e_i/s$$

define al *i*-ésimo residuo estandarizado donde s es la desviación estándar residual $(\sqrt{\text{CME}})$. Debe notarse que si el tamaño de la muestra n es muy grande, la distribución de los residuos estandarizados deberá encontrarse aproximada en forma adecuada por una distribución normal estándar. De hecho, muchos investigadores han sugerido que cualquier alejamiento notable de la normalidad en la distribución de los residuos puede indicar una deficiencia en el modelo.

Para determinar si un modelo de regresión es lineal o no en las variables de predicción, se grafican los residuos contra los correspondientes valores de cada una de las variables de predicción que figuran en la ecuación de regresión. Para determinar si la varianza del error es o no constante, se grafican los residuos estandarizados contra los correspondientes valores estimados de la respuesta. Finalmente, para determinar si una variable de predicción, potencialmente importante, debe incluirse o no en el modelo de regresión, se grafican los residuos contra los valores de esta variable. Si la ecuación de regresión estimada está prácticamente libre de cualquier deficiencia o violación de suposiciones, entonces los residuos estandarizados tenderán a encontrarse dentro de una banda horizontal centrada alrededor del valor cero, sin ninguna tendencia sistemática a ser positivos o negativos, y en forma muy rara se encontrarán fuera del intervalo ± 3. Cualquier desviación significativa con respecto a este comportamiento indicará la existencia de un problema.

La figura 14.2 representa algunas gráficas usuales de residuos: *a)* cuando se encuentra presente un efecto cuadrático causado por una variable de predicción y que debe incluirse en el modelo; *b)* cuando la varianza del error no es constante y deben emplearse mínimos cuadrados con factores de peso (ponderados) para estimar los coeficientes de regresión y *c)* cuando una variable que se ha eliminado muestra una fuerte asociación (lineal) con los residuos y por lo tanto debe incluirse en el modelo

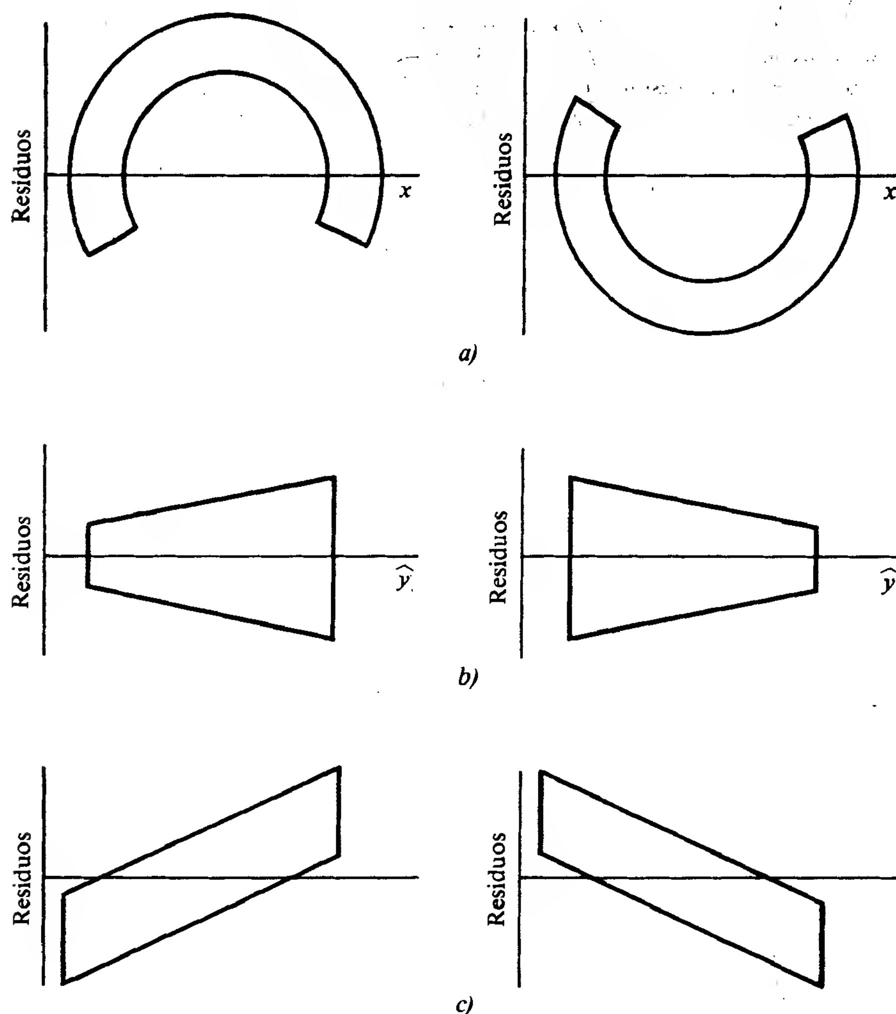


FIGURA 14.2 Gráficas comunes de residuos para: *a)* la presencia de un efecto cuadrático; *b)* la varianza no constante del error, y *c)* el efecto lineal de una variable omitida

de regresión. Puede decirse más con respecto a estos tres casos. Si la ecuación de regresión contiene sólo un efecto lineal causado por una variable de predicción x, cuando en realidad existe un efecto cuadrático estadísticamente apreciable, entonces la gráfica de los residuos estandarizados contra x será una curva en forma de U (o de U invertida). Bajo un efecto cuadrático, los residuos correspondientes a los valores extremos de x tenderán a ser grandes y positivos (negativos), y los residuos que se encuentran en la parte media del intervalo de valores de x tenderán a ser pequeños pero

negativos (positivos). En general, mediante la inclusión de un término cuadrático de x en el modelo, mejora considerablemente el valor predictivo de la ecuación de regresión resultante con respecto a la ecuación previa. Los efectos de orden superior también pueden detectarse de la misma manera.

Si la gráfica de los residuos da como resultado una figura en forma de cuña, entonces es posible que la suposición de que la varianza del error es constante no se cumpla. En otras palabras, si existe una tendencia a aumentar o disminuir los residuos estandarizados al aumentar los valores estimados de la respuesta, la varianza del error puede no ser constante. Esto da origen a lo que se conoce como *modelo heterocedástico*. Para remediar esta situación se emplea el método de *mínimos cuadrados con factores de peso,* en donde los pesos son inversamente proporcionales a la varianza de los errores. De esta forma, en lugar de intentar determinar las estimaciones de los coeficientes de regresión mediante la minimización de la suma de los cuadrados de los errores, se determina el conjunto de valores para los cuales la suma de pesos de los cuadrados de los errores es un mínimo. El motivo para emplear mínimos cuadrados con factores de peso en una situación heterocedástica es estimar los coeficientes de regresión con pequeñas desviaciones lo que a su vez produce un mejor ajuste.

Si cuando los residuos estandarizados se grafican contra una variable que no forma parte de la ecuación de regresión, pero bajo la cual se pudo observar la respuesta, se observa una tendencia lineal (o de orden superior); entonces, como se mencionó en el capítulo 13, los errores no pueden considerarse más como independientes de esta variable. En general, este tipo de variable resulta ser un efecto demográfico o relacionado con el tiempo. Por ejemplo, para muchos experimentos en los que los datos se observan durante un periodo significativo, el investigador podría inicialmente decidir no incluir al tiempo como una variable de predicción potencial. Pero si los residuos revelan un patrón sistemático cuando se grafican contra el tiempo, la variable tiempo deberá introducirse en la ecuación de regresión.

Las gráficas de residuos también son una ayuda al tratar con observaciones extremas o discrepantes. En general, las observaciones extremas tienen residuos que son, en forma relativa, grandes, comparados con los de las demás observaciones. En general, el valor del residuo estandarizado de una observación discrepante se encontrará más allá del intervalo ± 3. Las observaciones discrepantes pueden crear situaciones difíciles en una ecuación de regresión, debido a que tienen un efecto desproporcionado sobre los valores estimados de los coeficientes de regresión. Recuérde que una de las suposiciones de la estimación por mínimos cuadrados es que el conjunto de datos es típico de la situación para la cual se intenta identificar una buena ecuación de predicción. Por lo tanto, la remoción de cualquier observación del conjunto de datos no tendrá, en forma virtual, ningún efecto sobre la ecuación de regresión. Lo anterior constituye precisamente el porqué puede removerse, sólo con extremo cuidado, una observación discrepante. Un método lógico que se ha sugerido es remover una observación discrepante sólo si existe evidencia comprobada de que ésta es el producto de un error causado, por ejemplo, por un mal funcionamiento del instrumento de medición. En ausencia de clara evidencia de error, la observación discrepante puede ser información única con respecto a la respuesta y ser vital para el entendimiento del fenómeno.

Los siguientes dos ejemplos ilustrarán los casos *a)* y *c)* que se muestran en la figura 14.2. El caso en el cual se tiene una varianza no constante se analizará en la sección 14.8.

Ejemplo 14.3 Una compañía manufacturera desea predecir el costo unitario de fabricación Y de uno de sus productos como una función de la tasa de producción (que fluctúa en el tiempo) x_1 y de los costos de material y mano de obra x_2. Los datos se recabaron durante un periodo de 20 meses durante el cual la tasa de producción y los costos del material y la mano de obra experimentaron una fluctuación muy amplia. La tasa de producción se midió como un porcentaje de la capacidad total de producción, y se utilizó un índice apropiado para reflejar los costos del material y mano de obra. Las observaciones se encuentran en la tabla 14.11. Obténgase la mejor ecuación de regresión para predecir el costo por unidad.

Primero se supondrá un modelo de regresión lineal que sólo tome en cuenta a x_1 y a x_2. En la tabla 14.12 se proporcionan las estimaciones y otra información importante. Hasta aquí parece que todo marcha muy bien. Las estimaciones tienen sentido (valor negativo para el coeficiente x_1 y positivo para el de x_2), las desviaciones estándar son pequeñas, el valor de R^2 es relativamente alto y todos los efectos son estadísticamente discernibles. Por lo tanto, se podría concluir que se ha obtenido una buena ecuación de predicción, pero una gráfica de los residuos estandarizados contra x_1 revela un patrón cuadrático en la mitad superior de la figura 14.3. Ningún patrón es evidente para x_2 .

TABLA 14.11 Datos de la muestra para el ejemplo 14.3

Y	x_1	x_2
13.59	87	80
15.71	78	95
15.97	81	106
20.21	65	115
24.64	51	128
21.25	62	128
18.94	70	115
14.85	91	92
15.18	94	93
16.30	100	111
15.93	102	116
16.45	82	117
19.02	74	127
18.16	85	133
18.57	86	135
17.01	90	136
18.03	93	140
19.22	81	142
21.12	72	148
23.32	60	150

TABLA 14.12 Análisis de regresión para el ejemplo 14.3

Regresión de Y sobre x_1 y x_2

Variable	*Coeficiente de regresión estimado*	*Desviación estándar estimada*	*Valor t*
Constante	20.2800	2.1300	9.54
x_1	−0.1377	0.0159	−8.69
x_2	0.0742	0.0110	6.77

$R^2 = 0.914$ $\qquad t_{0.975,\,17} = 2.11$

ANOVA

Fuente	gl	SC	CM	*Valor F*
Regresión	2	144.39	72.19	90.24
x_1	1	107.72	107.72	134.65
$x_2 \mid x_1$	1	36.67	36.67	45.84
Error	17	13.59	0.80	
Total	19	157.98	$f_{0.95,\,2,\,17} = 3.59$; $f_{0.95,\,1,\,17} = 4.45$	

La gráfica de los residuos para x_1 implica que debe incluirse un término cuadrático en x_1 en el modelo de regresión. De esta forma, se ajustará el modelo

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \varepsilon$$

obteniéndose los resultados que se muestran en la tabla 14.13.

Una comparación con los resultados anteriores revela que la inclusión de un efecto cuadrático en x_1 mejora en forma considerable la ecuación de regresión estimada. Por ejemplo, los coeficientes de regresión, tanto de x_1 como de x_2, se estiman con una mejor precisión comparada con la anterior y el valor de R^2 se incrementa hasta 0.981. Además, la nueva gráfica de residuos contra x_1 (véase la Fig. 14.4) no muestra ningún patrón apreciable.

Ejemplo 14.4 Recuérdese el ejemplo de los salarios iniciales contra la calificación promedio, empleado a través de todo el capítulo 13. Quizá el lector se pregunte si existiesen otras variables de predicción potenciales. Supóngase que también se ha observado la edad de cada estudiante en la muestra. Ya que algunas compañías tienen como requisito poseer alguna experiencia en el campo y un recién egresado de mayor edad podría tenerla, es posible que la edad de éste pueda influenciar en el salario inicial que percibirá. Los datos, tomando en cuenta la edad, se encuentran en la tabla 14.14.

Cuando se hace una gráfica de los residuos estandarizados de la ecuación de regresión estimada $\hat{y} = -6.63 + 8.12x_1$ contra los correspondientes valores de x_2

TABLA 14.13 Análisis de regresión revisado para el ejemplo 14.3

Regresión de Y sobre x_1 y x_2 y x_1^2

Variable	*Coeficiente de regresión estimado*	*Desviación estándar estimada*	*Valor t*
Constante	41.550000	3.050000	13.64
x_1	−0.700300	0.076200	−9.20
x_2	0.073400	0.005400	13.68
x_1^2	0.003624	0.000488	7.43

$R^2 = 0.981$ $\qquad t_{0.975,\,16} = 2.12$

ANOVA

Fuente	gl	SC	CM	*Valor F*
Regresión	3	154.92	51.640	270.37
x_1	1	107.72	107.72	563.98
$x_2 \mid x_1$	1	36.66	36.66	191.94
$x_1^2 \mid x_1, x_2$	1	10.54	10.54	55.18
Error	16	3.06	0.191	
Total	19	157.98	$f_{0.95,\,3,\,16} = 3.24$; $f_{0.95,\,1,\,16} = 4.49$	

(véase la Fig. 14.5), se observa una tendencia lineal ascendente. Por lo tanto, se incluye el efecto lineal de x_2 en el modelo de regresión y se ajusta

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon.$$

En la tabla 14.15 se muestran los nuevos resultados. Dado que ahora se estiman con mejor precisión el término constante, el coeficiente de x_1 y el valor de R^2 ha aumentado en forma apreciable, la inclusión de x_2 da como resultado una mejor ecuación de predicción.

14.7 Regresión polinomial

En la sección 14.2 se mencionó que el modelo polinomial dado por (14.3), o alguno que contenga términos de interacción como (14.4), es un caso especial del modelo lineal general. De hecho, en el ejemplo 14.3 se mostró cómo el efecto cuadrático de una variable de predicción puede mejorar la capacidad predictiva de la ecuación de regresión. En esta sección se ahondará más sobre este tipo de modelos.

Si se ha identificado sólo una variable de predicción x y la gráfica de las respuestas observadas contra los valores de x revela una curvatura, entonces debe usarse un polinomio en x, de cierto grado, para aproximar la verdadera curva de regresión.

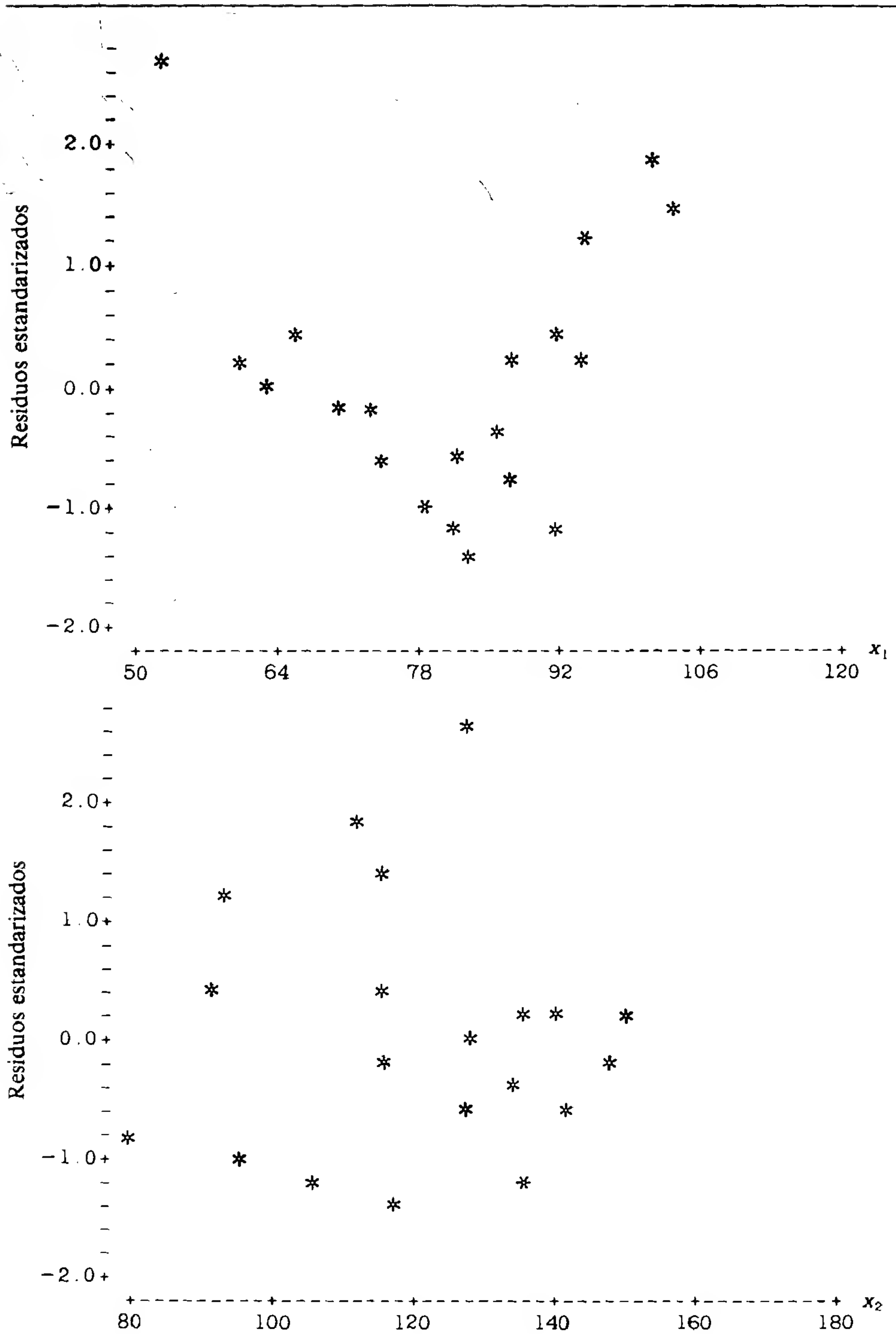


FIGURA 14.3 Gráficas de los residuos estandarizados para el ejemplo 14.3

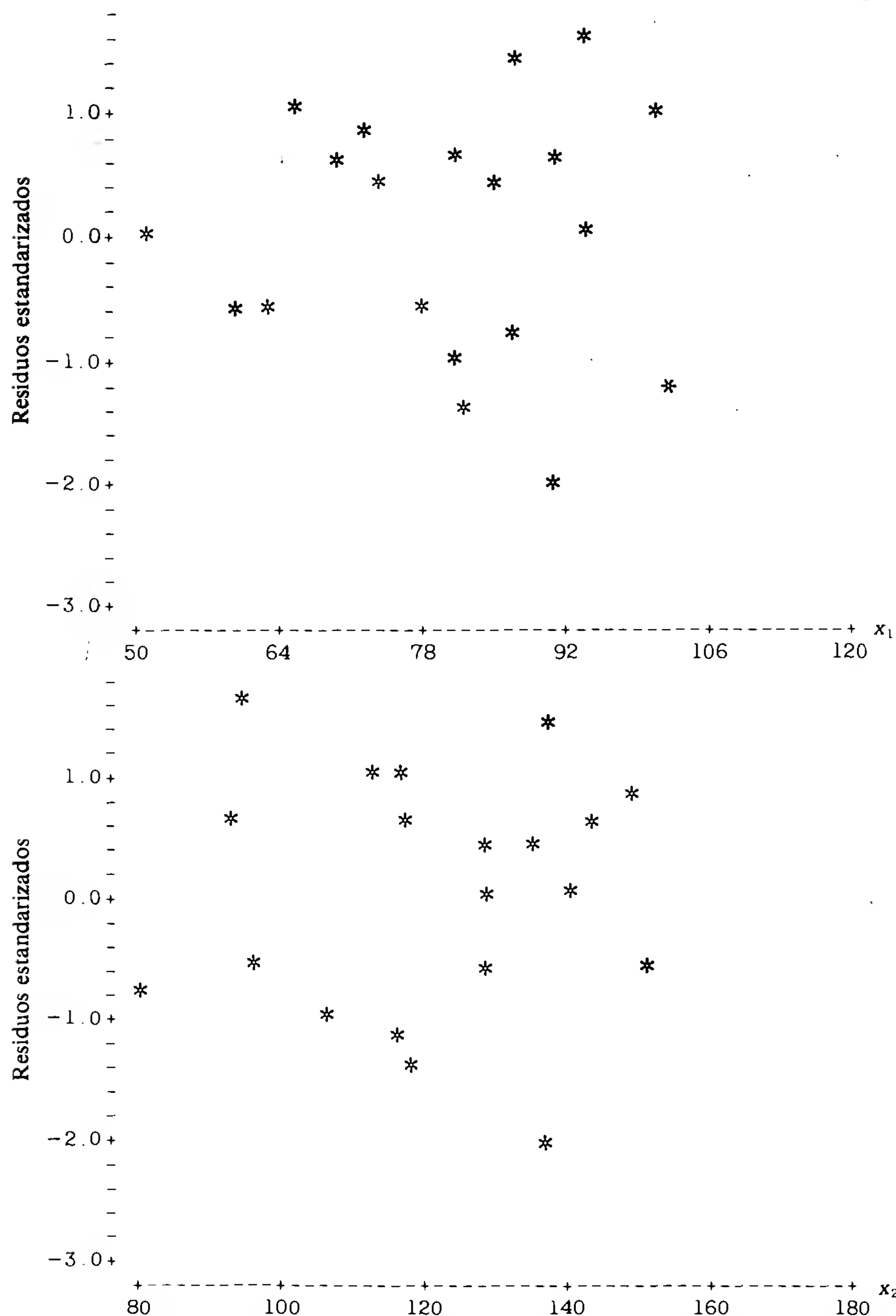


FIGURA 14.4 Gráficas de los residuos estandarizados para la ecuación de regresión revisada en el ejemplo 14.3

TABLA 14.14 Datos aumentados para el ejemplo de los salarios iniciales

Y	x_1 (*CP*)	x_2 *Edad*
18.5	2.95	22
20.0	3.20	23
21.1	3.40	23
22.4	3.60	23
21.2	3.20	27
15.0	2.85	22
18.0	3.10	25
18.8	2.85	28
15.7	3.05	23
14.4	2.70	22
15.5	2.75	28
17.2	3.10	22
19.0	3.15	26
17.2	2.95	23
16.8	2.75	26

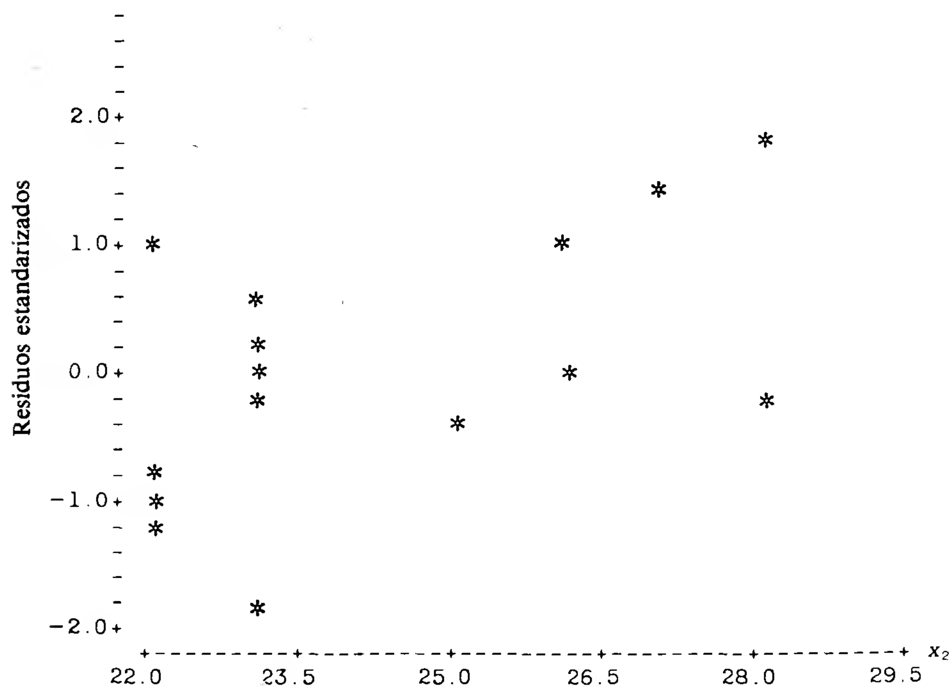


FIGURA 14.5 Residuos estandarizados contra la edad para el ejemplo 14.4

TABLA 14.15 Análisis de regresión para el ejemplo 14.4

Variable	*Coeficiente de regresión estimado*	*Desviación estándar estimada*	*Valor t*
Constante	−16.880	5.470	−3.05
x_1	8.740	1.220	7.16
x_2	0.338	0.137	2.47

$R^2 = 0.813$ $\qquad t_{0.975,\,12} = 2.179$

ANOVA

Fuente	gl	SC	CM	*Valor F*
Regresión	2	66.10	33.05	26.23
x_1	1	58.40	58.40	46.35
$x_2 \mid x_1$	1	7.70	7.70	6.11
Error	12	15.17	1.26	
Total	14	81.27	$f_{0.95,\,2,\,12} = 3.89$; $f_{0.95,\,1,\,12} = 4.75$	

Por ejemplo, un modelo cúbico en x está dado por

$$Y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \beta_{111} x_i^3 + \varepsilon_i,$$

donde β_1 recibe el nombre de *coeficiente lineal,* β_{11} es el *coeficiente cuadrático* y β_{111} es el *coeficiente cúbico.* Para mantener la costumbre se ha alterado en forma ligera la notación para estos coeficientes de regresión para reflejar el patrón de la correspondiente potencia de x.

Como se mencionó con anterioridad, lo que se busca con un polinomio es el grado que mejor ajuste los datos dados. De acuerdo con lo anterior, el interés recae en probar hipótesis, como por ejemplo, H_0: $\beta_{11} = 0$ o H_0: $\beta_{111} = 0$. Mediante el empleo de este enfoque se tiene la capacidad para determinar el polinomio más apropiado para estimar la respuesta promedio. Sin embargo, se advierte al lector que lo que se busca y se prefiere en forma general es un polinomio de un orden relativamente bajo. Se deberá evitar el empleo de potencias muy grandes de la variable de predicción, debido a que lo que ocurre la mayor parte de las veces es un ajuste que explica incluso las variaciones aleatorias que se encuentran en los datos; en otras palabras, siempre se puede encontrar un modelo polinomial de un grado, lo suficientemente alto para ajustar los datos de manera perfecta, ya que un polinomio de grado $n - 1$ pasará a través de todos los n valores de la respuesta.

Muchas veces un modelo completo de segundo orden que contiene términos lineales, cuadráticos y de interacción, proporciona una aproximación funcional excelente en comparación con una función de respuesta desconocida y, en forma general, compleja. Por ejemplo, un modelo de segundo orden en dos variables es

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} + \varepsilon_i,$$

donde β_1 y β_2 son los coeficientes lineales, β_{11} y β_{22} son los coeficientes cuadráticos y β_{12} es el coeficiente de interacción. Para este modelo, la matriz **X** y el vector de parámetros $\boldsymbol{\beta}$ que figuran en la ecuación matricial

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

son

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{11}^2 & x_{12}^2 & x_{11}x_{12} \\ 1 & x_{21} & x_{22} & x_{21}^2 & x_{22}^2 & x_{21}x_{22} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n1}^2 & x_{n2}^2 & x_{n1}x_{n2} \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{12} \end{bmatrix}.$$

Con los dos siguientes ejemplos se ilustrarán tanto un modelo polinómico en una variable, así como un modelo completo de segundo orden.

Ejemplo 14.5 La demanda de cierto producto cambió debido a una variación rápida de su precio por unidad. Supóngase que la demanda Y del producto se observa en una región geográfica en particular sobre un intervalo bastante amplio de precios x. Dados los datos que se encuentran en la tabla 14.16, determínese el grado de un polinomio que mejor ajuste estos datos.

Dado que sólo se tiene una variable de predicción, lo primero que se tiene que hacer es una gráfica de la demanda contra el precio por unidad. La figura 14.6 revela una curvatura, lo cual indica que debe intentarse el ajuste con un modelo cuadrático.

Para ilustrar cómo se detecta la curvatura, supóngase que se propone un modelo lineal sencillo. En la figura 14.7 se muestra un listado de computadora generado por Minitab y los residuos estandarizados resultantes contra el precio en la figura 14.8. La necesidad de incluir un efecto cuadrático en x es evidente.

TABLA 14.16 Datos de la muestra para el ejemplo 14.5

Y unidades	*x dólares*
360	8.8
305	9.7
230	9.9
242	10.3
180	11.0
172	12.5
121	13.2
83	14.8
122	15.8
91	17.4
105	18.2

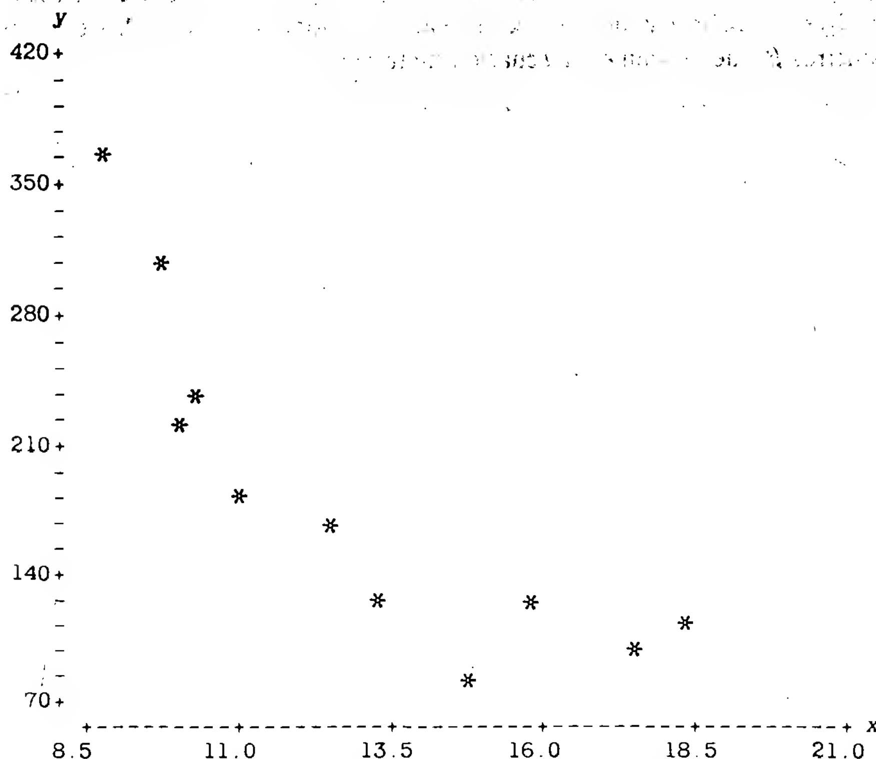


FIGURA 14.6 Gráfica de la demanda contra el precio por unidad para el ejemplo 14.5

LA ECUACION DE REGRESION ES
Y = 497. − 24.4 X1

	COLUMNA	COEFICIENTE	DEV. EST. DEL COEF.	COEFICIENTE-T =COEF/D.E.
	—	497.15	60.85	8.17
X1	C2	−24.419	4.594	−5.32

LA DEV. EST. DE Y CON RESPECTO A LA RECTA DE REGRESION ES
S = 47.53
CON (11− 2) = 9 GRADOS DE LIBERTAD

R-CUADRADA = 75.8 POR CIENTO

ANALISIS DE VARIANZA

DEBIDO A	GL	SC	CM = SC/GL
REGRESION	1	63815	63815
RESIDUO	9	20330	2259
TOTAL	10	84145	

FIGURA 14.7 Listado de computadora para el ejemplo 14.5

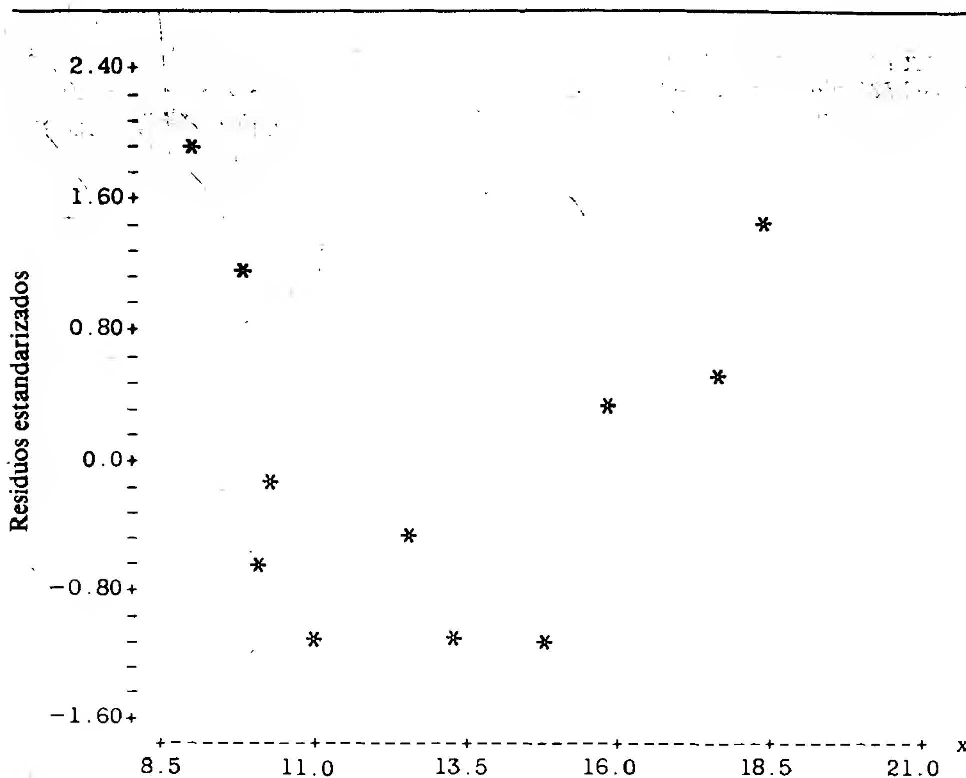


FIGURA 14.8 Residuos estandarizados contra el precio por unidad para el modelo lineal del ejemplo 14.5

El listado de salida para un modelo cuadrático se muestra en la figura 14.9. Como se esperaba, existe una considerable mejoría en la predicción proporcionada por la ecuación de regresión estimada, que la que se tenía con un modelo lineal simple. Nótese que Minitab también proporciona la "SC tipo *I*", es decir, a través de las entradas identificadas por *"C2"* y *"C3"* se tiene que SCR(x) = 63 814.5 y SCR($x^2 \mid x$) = 14 961.4, respectivamente.

Aunque no se proporciona una gráfica de los residuos estandarizados contra el precio para el modelo cuadrático, no mostrará ningún patrón evidente; además, no se obtiene ninguna mejoría apreciable si se añaden al modelo términos de orden superior. Una ecuación de regresión estimada de orden cuadrático es lo más adecuado para predecir la demanda de este producto como una función del precio por unidad.

Ejemplo 14.6 En el ejemplo 14.2 se consideró la regresión lineal de la temperatura aparente Y sobre la temperatura del aire x_1 y la humanidad relativa x_2 para un intervalo limitado de x_1 y x_2. Para el conjunto aumentado de datos dado en la tabla 14.17 se desea ajustar y analizar una ecuación de regresión completa de segundo orden.

```
LA ECUACION DE REGRESION ES
Y =  1330. - 155. X1 + 4.87 X2
                                                    DEV. EST.    COCIENTE-T
         COLUMNA            COEFICIENTE             DEL COEF.    = COEF/D.E.
          —                      1330.4               179.6          7.41
X1        C2                   -155.47                27.87         -5.58
X2        C3                     4.866                1.031          4.72

LA DEV. EST. DE Y CON RESPECTO A LA RECTA DE REGRESION ES
S = 25.91
CON  (  11- 3) =  8 GRADOS DE LIBERTAD

R-CUADRADA = 93.6 POR CIENTO

ANALISIS DE VARIANZA

 DEBIDO A        GL            SC          CM = SC/GL
REGRESION         2       78775.8            39387.9
RESIDUO           8        5368.8              671.1
TOTAL            10       84144.7

ANALISIS DE VARIANZA ADICIONAL
SC EXPLICADA POR CADA VARIABLE QUE ENTRE EN EL ORDEN DADO

 DEBIDO A        GL            SC
REGRESION         2       78775.8
C2                1       63814.5
C3                1       14961.4
```

FIGURA 14.9 Listado revisado para el ejemplo 14.5

TABLA 14.17 Datos de la muestra para el ejemplo 14.6

x_2 \ x_1	70°	75	80	85	90	95
0%	64	69	73	78	83	87
10	65	70	75	80	85	90
20	66	72	77	82	87	93
30	67	73	78	84	90	96
40	68	74	79	86	93	101
50	69	75	81	88	96	107
60	70	76	82	90	100	114
70	70	77	85	93	106	124
80	71	78	86	97	113	136

Con base en la experiencia cotidiana de cualquier persona con respecto al clima, debe ser evidente que la temperatura del aire y la humedad relativa tienen una interacción con la temperatura aparente. Por ejemplo, el calor que se siente cuando la temperatura del aire es de 90° y la humedad relativa es del 10%, es muy diferente a la que se percibe cuando la humedad relativa es del 70%. Los resultados que se muestran en la tabla 14.18 son los que se obtienen cuando se supone un modelo completo de segundo orden.

El efecto de cada término en el modelo sobre la temperatura aparente es estadísticamente discernible; los coeficientes de regresión se encuentran estimados con una exactitud razonablemente buena y el valor de R^2 es muy alto. De esta forma, la ecuación de regresión estimada completa de segundo orden es adecuada para la predicción.

14.8 Mínimos cuadrados con factores de peso

Una suposición clave en la estimación por mínimos cuadrados es que la varianza de cada error aleatorio es la misma. De la sección 14.6 recuérdese que si los residuos es-

TABLA 14.18 Análisis de regresión para el ejemplo 14.6

Regresión de Y sobre x_1, x_2, x_1^2, x_2^2, y x_1x_2

Variable	*Coeficiente de regresión estimado*	*Desviación estándar estimada*	*Valor t*
Constante	175.3300	36.11000	4.86
x_1	−3.1689	0.87580	−3.62
x_2	−1.4351	0.13210	−10.87
x_1^2	0.0236	0.00530	4.46
x_2^2	0.0017	0.00056	3.07
x_1x_2	0.0188	0.00150	12.56
	$R^2 = 0.977$	$t_{0.975,48} = 2.01$	

ANOVA

Fuente	gl	SC	CM	*Valor F*
Regresión	5	11,966.71	2393.34	407.20
Efecto lineal de x_1	1	8536.13	8536.13	1452.32
Efecto lineal de x_2	1	2330.71	2330.71	396.54
$x_1^2 \mid x_1, x_2$	1	116.68	116.68	19.85
$x_2^2 \mid x_1, x_2, x_1^2$	1	55.41	55.41	9.43
Interacción de x_1, x_2	1	927.78	927.78	157.85
Error	48	282.12	5.88	
Total	53	12,248.83	$f_{0.95,5,48} = 2.42$; $f_{0.95,1,48} = 4.04$	

tandarizados tienden a disminuir o a aumentar conforme se incrementan los valores estimados de la respuesta, la varianza del error no puede considerarse como constante. El remedio apropiado para esta situación es aplicar mínimos cuadrados con *factores de peso*, en los cuales las estimaciones para los coeficientes de regresión se obtienen mediante la minimización de la suma con pesos de los cuadrados de los errores. Si se empleara la estimación por mínimos cuadrados ordinarios en una situación para la cual la varianza del error no es constante, los coeficientes de regresión no serían estimados con la misma precisión.

Antes de resolver algunos ejemplos, se examinarán en forma breve los aspectos teóricos clave de la estimación por mínimos cuadrados con factores de peso. Al igual que en los mínimos cuadrados ordinarios se supone que para el modelo lineal general

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

$\boldsymbol{\varepsilon}$ es un vector de errores aleatorios no observable, tal que

$$E(\boldsymbol{\varepsilon}) = \mathbf{0},$$

y la matriz de varianza-covarianza está dada por

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \mathbf{Q}.$$

La matriz $\mathbf{Q}$ es de tal naturaleza que el elemento que se encuentra sobre la diagonal q_{ii} es la varianza de ε_i, y q_{ij} es la covarianza entre ε_i y ε_j para toda $i \neq j$. $\mathbf{Q}$ debe ser no singular; de hecho, $\mathbf{Q}^{-1}$ recibe el nombre de matriz de ponderación y la debe especificar el investigador, es decir, los pesos se asignan a cada observación de la respuesta de acuerdo con alguna información respecto a la correspondiente varianza del error. Existen algunos procedimientos disponibles para los usuarios para determinar los pesos; lo anterior se ilustrará más adelante.

Las estimaciones de los coeficientes de regresión se obtienen mediante la minimización de la suma con pesos de los cuadrados de los errores dada por

$$\boldsymbol{\varepsilon}'\mathbf{Q}^{-1}\boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{Q}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}).$$

Puede demostrarse que las ecuaciones normales en forma matricial son

$$\mathbf{X}'\mathbf{Q}^{-1}\mathbf{X}\mathbf{B} = \mathbf{X}'\mathbf{Q}^{-1}\mathbf{Y}.$$

Si existe la matriz inversa $(\mathbf{X}'\mathbf{Q}^{-1}\mathbf{X})^{-1}$, los estimadores por mínimos cuadrados con factores de peso se obtienen mediante

$$\mathbf{B} = (\mathbf{X}'\mathbf{Q}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}^{-1}\mathbf{Y}. \qquad (14.31)$$

Es importante notar que los mínimos cuadrados ordinarios son un caso especial de los mínimos cuadrados con factores de peso, es decir, si $\mathbf{Q} = \sigma^2\mathbf{I}$, entonces es relativamente fácil demostrar que (14.31) se reduce a la expresión usual

$$\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

La definición de la matriz **Q** implica una estructura de covarianza entre los errores aleatorios. En la práctica, esta estructura resulta difícil de identificar. La aplicación más sencilla de la estimación por mínimos cuadrados con factores de peso es la de suponer que **Q** es una matriz diagonal de la forma

$$\mathbf{Q} = \begin{bmatrix} \sigma_1^2 & & & & \\ & \sigma_2^2 & & 0 & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & & \cdot \\ & & & & \sigma_n^2 \end{bmatrix},$$

donde σ_i^2 es la varianza de ε_i. Entonces

$$\mathbf{Q}^{-1} = \begin{bmatrix} 1/\sigma_1^2 & & & & \\ & 1/\sigma_2^2 & & 0 & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & & \cdot \\ & & & & 1/\sigma_n^2 \end{bmatrix}.$$

Por lo tanto, los errores aleatorios se suponen independientes, pero algunas de sus varianzas (si no es que todas) pueden ser diferentes.

A continuación se examinarán algunas situaciones para las cuales es probable que se viole la suposición de que la varianza del error es constante si se emplean mínimos cuadrados ordinarios. Una práctica muy frecuente en la adquisición de datos experimentales es tomar varias mediciones de la respuesta para cada uno de los puntos de observación y después calcular el promedio de las mediciones para cada uno. La principal razón para llevar a cabo este procedimiento es estabilizar la variabilidad de las observaciones individuales. Bajo este procedimiento la respuesta se convierte en un promedio. Dado que la desviación estándar de un promedio es proporcional a la raíz cuadrada del tamaño de la muestra sobre la cual se basa este promedio, la variación de $\overline{Y}_i$, y de esta forma de ε_i, es σ^2/n_i, donde σ^2 es la varianza común del error y n_i es el tamaño de la muestra en relación con $\overline{Y}_i$. Esto conduce a un procedimiento de estimación por mínimos cuadrados con factores de peso para el cual la inversa de la matriz **Q** está dada por

$$\mathbf{Q}^{-1} = \frac{1}{\sigma^2}\begin{bmatrix} n_1 & & & & \\ & n_2 & & 0 & \\ & & \cdot & & \\ & & & \cdot & \\ & 0 & & & \cdot \\ & & & & n_n \end{bmatrix}.$$

Los pesos son los tamaños individuales de cada muestra $n_1, n_2, \ldots, n_n$ para los n puntos de observación. La lógica que se encuentra detrás de lo anterior es muy senci-

lla; los promedio basados en un gran número de observaciones deben tener un mayor peso en la determinación de las estimaciones que aquellas que se basan en pocas observaciones.

Ejemplo 14.7 Una compañía implanta un programa de inspección en el cual las unidades de cierto producto se revisan en forma visual en la línea de producción, con el objetivo de detectar las que se encuentran defectuosas. El gerente sabe que la velocidad de la línea afectará el número de unidades defectuosas encontradas. Se selecciona un lote de unidades de un tamaño suficiente y se envía a un departamento que se encargará de revisar el 100% de los mismos minuciosamente con el propósito de encontrar el número total de unidades defectuosas. Entonces el lote se coloca sobre la línea un número variable de veces para cada una de las ocho velocidades que posee la misma. Para cada velocidad de la línea x se calcula el número promedio $\overline{Y}$ de unidades defectuosas que no se descubrieron. Los datos que aparecen en la tabla 14.19 son los resultados de este experimento y la última columna representa los tamaños individuales de cada muestra. Obténganse y compárense las regresiones simples de $\overline{Y}$ sobre x, con base en mínimos cuadrados ordinarios y con factores de peso.

Para mínimos cuadrados ordinarios se descartan los tamaños variables de cada muestra y se procede de la manera usual. Los resultados se muestran en la tabla 14.20. Para mínimos cuadrados con factores de peso se ilustrará el cálculo de las estimaciones. Dado que los pesos son los tamaños de cada una de las muestras,

$$\mathbf{Q}^{-1} = \frac{1}{\sigma^2}\begin{bmatrix} 14 & & & & 0 \\ & 3 & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & & 2 \end{bmatrix},$$

TABLA 14.19 Datos de la muestra para el ejemplo 14.7

$\overline{Y}$	x(pie/min)	n
0.50	10	14
4.67	20	3
6.25	30	25
10.00	40	2
13.50	50	3
13.70	60	22
17.50	70	5
23.00	80	2

TABLA 14.20 Estimaciones por mínimos cuadrados ordinarios y tabla ANOVA para el ejemplo 14.7

Variable	*Coeficiente de regresión estimado*	*Desviación estándar estimada*	*Valor t*
Constante	−2.1190	0.9490	−2.23
x	0.2946	0.0188	15.68
	$r^2 = 0.976$	$t_{0.975,\,6} = 2.447$	

ANOVA

Fuente	gl	SC	CM	*Valor F*
Regresión	1	364.62	364.62	246.36
Error	6	8.89	1.48	
Total	7	373.51	$f_{0.95,\,1,\,6} = 5.99$	

y

$$\mathbf{X'Q^{-1}X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 10 & 20 & \cdots & 80 \end{bmatrix} \frac{1}{\sigma^2} \begin{bmatrix} 14 & & & & \\ & 3 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & 2 \end{bmatrix} \begin{bmatrix} 1 & 10 \\ 1 & 20 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & 80 \end{bmatrix}$$

$$= \frac{1}{\sigma^2} \begin{bmatrix} 76 & 3010 \\ 3010 & 152\,300 \end{bmatrix}$$

Además,

$$(\mathbf{X'Q^{-1}X})^{-1} = \sigma^2 \begin{bmatrix} 0.06056388 & -0.00119696 \\ -0.00119696 & 0.00003022 \end{bmatrix}$$

Entonces, mediante el empleo de (14.31), las estimaciones de mínimos cuadrados con factores de peso son

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \sigma^2 \begin{bmatrix} 0.06056388 & -0.00119696 \\ -0.00119696 & 0.00003022 \end{bmatrix} \begin{bmatrix} 1 & 1 \cdots & 1 \\ 10 & 20 \cdots & 80 \end{bmatrix}$$

$$\frac{1}{\sigma^2}\begin{bmatrix} 14 & & & 0 \\ & 3 & & \\ & & \ddots & \\ 0 & & & 2 \end{bmatrix} \cdot \begin{bmatrix} 0.5 \\ 4.67 \\ \cdot \\ \cdot \\ \cdot \\ 23.0 \end{bmatrix} = \begin{bmatrix} -2.0540 \\ 0.2753 \end{bmatrix}.$$

Los resultados del análisis de regresión con factores de peso se dan en la tabla 14.21.

Una comparación de los resultados basados en mínimos cuadrados con factores de peso y ordinarios muestran una ligera mejoría en la precisión de las estimaciones de mínimos cuadrados con factores de peso, así como un pequeño aumento en el valor de r^2. Debe notarse que si los tamaños de cada muestra individual no difieren mucho entre sí, es muy probable que los resultados que se obtengan mediante el empleo de los dos métodos sean muy similares.

En el ejemplo 14.7 se supuso de antemano una varianza del error no constante, debido a que el registro de las observaciones de la respuesta son promedios basados en tamaños variables de las muestras. Sin embargo, la mayoría de las veces la falta de una varianza constante para el error debe determinarse en forma empírica. Ya se ha indicado que una gráfica de los residuos estandarizados contra las respuestas estimadas resulta ser una ayuda considerable en la detección de la heterocedasticidad, pero para aquellos problemas para los cuales se tienen disponibles observaciones repetidas de la respuesta para el mismo punto de observación, es preferible registrar las observaciones reales más que sus promedios y emplearlas para detectar una varianza

TABLA 14.21 Estimaciones por mínimos cuadrados con factores de peso y tabla ANOVA para el ejemplo 14.7

Variable	*Coeficiente de regresión estimado*	*Desviación estándar estimada*	*Valor t*
Constante	−2.0540	0.6990	−2.94
x	0.2753	0.0156	17.63
	$r^2 = 0.981$	$t_{0.975,\,6} = 2.447$	

ANOVA

Fuente	gl	SC	CM	*Valor F*
Regresión	1	2508.66	2508.66	310.86
Error	6	48.43	8.07	
Total	7	2557.09*	$f_{0.95,\,1,\,6} = 5.99$	

* Las sumas de los cuadrados son diferentes de las anteriores debido a que ahora son funciones de los pesos impuestos.

no constante del error. Para estos tipos de problemas la gráfica de las observaciones contra los valores de la variable de predicción revelará una varianza no constante, si es que ésta existe. El siguiente ejemplo ilustra este problema.

Ejemplo 14.8 Recientemente, la variabilidad del ozono en la estratósfera ha recibido gran atención, especialmente en el impacto que el hombre tiene sobre el clima. El ozono es una forma de oxígeno que se encuentra en diversas cantidades en la estratósfera y constituye un componente muy importante de la atmósfera, ya que tiene la propiedad de bloquear la radiación ultravioleta que provienen del sol. Los datos que se encuentran en la tabla 14.22 muestran la cantidad de ozono registrada Y y su presión parcial x para cada capa de altitud, donde cada capa tiene aproximadamente un kilómetro de altura. Por conveniencia, las capas se han escalado a un intervalo de -7 a $+7$. Determínese si la varianza del error puede considerarse como constante.

TABLA 14.22 Datos de la muestra para el ejemplo 14.8

Capa	*Ozono*	*Capa*	*Ozono*
−7.00	53.8	−1.00	102.8
−7.00	53.3	−1.00	96.9
−7.00	54.8	−1.00	98.2
−7.00	54.6	0.0	98.9
−7.00	53.7	0.0	96.1
−7.00	55.2	0.0	99.6
−7.00	55.7	0.0	91.4
−7.00	54.1	1.00	101.1
−6.00	63.8	1.00	94.6
−6.00	64.2	1.00	95.9
−6.00	66.9	2.00	92.3
−6.00	67.2	2.00	96.6
−6.00	65.4	2.00	98.5
−6.00	67.3	3.00	93.6
−5.00	71.8	3.00	86.2
−5.00	73.2	3.00	87.9
−5.00	75.6	3.00	89.5
−5.00	76.2	4.00	74.8
−5.00	72.7	4.00	82.3
−4.00	79.4	4.00	76.9
−4.00	81.1	4.00	81.2
−4.00	85.2	5.00	73.6
−4.00	83.0	5.00	65.4
−4.00	84.1	5.00	67.1
−4.00	82.8	6.00	60.2
−3.00	90.3	6.00	54.9
−3.00	84.2	6.00	50.8
−3.00	88.3	7.00	44.7
−3.00	86.0	7.00	38.5
−2.00	93.2		
−2.00	97.4		
−2.00	98.3		

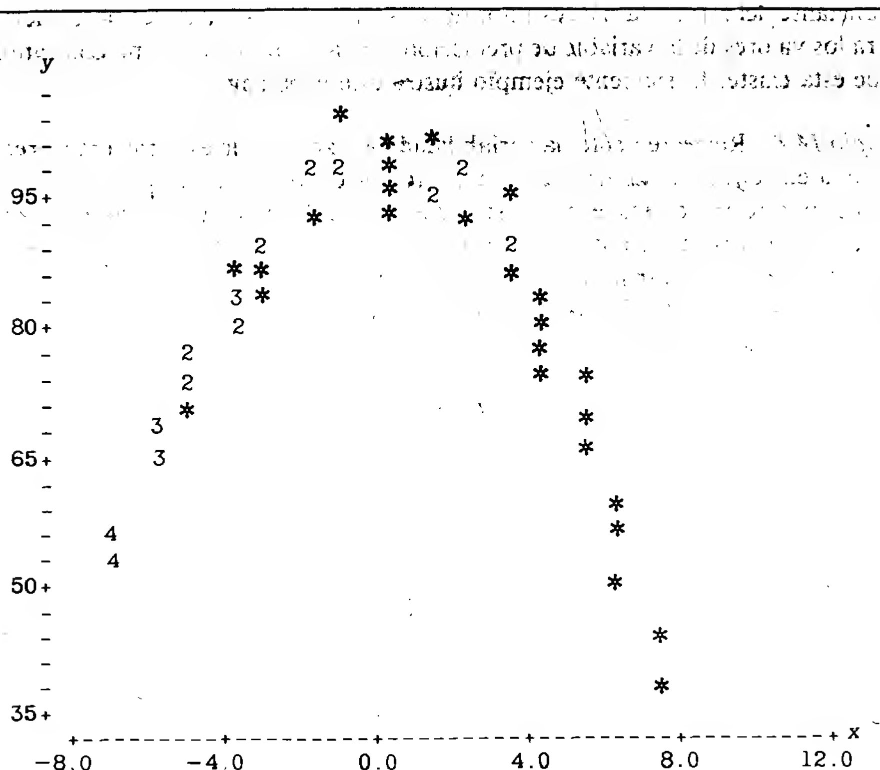


FIGURA 14.10 Gráfica del ozono contra la altitud de la capa para el ejemplo 14.8

Una gráfica de la cantidad de ozono contra la capa, figura 14.10, revela que la varianza del error no puede considerarse como constante debido a que la variabilidad en la cantidad de ozono aumenta conforme la capa crece. La figura 14.10 también sugiere que el modelo apropiado por utilizar es una ecuación cuadrática.

En una situación como ésta, en la que existen repeticiones para varios puntos de observación, los pesos se determinan mediante el cálculo de la varianza de las mediciones de la respuesta para cada punto de observación. De esta forma, cada peso es el recíproco de la correspondiente varianza. Por ejemplo, si y_{ij} denota la i-ésima medición de ozono en la j-ésima capa, la varianza de la muestra de la j-ésima capa es

$$s_j^2 = \sum_{j=1}^{n_j} (y_{ij} - \bar{y}_{\cdot j})^2/(n_j - 1),$$

y el correspondiente peso es $w_j = 1/s_{\cdot j}^2$. Como ilustración, considérense las observaciones para $x = 0$. Éstas son 98.9, 99.6 y 91.4. Entonces, $n_j = 4$, $\bar{y}_{\cdot j} = 96.5$, $s_j^2 = 13.8467$, y $w_j = 1/13.8467 = 0.0722$. Al seguir este procedimiento, los pesos correspondientes para cada capa son los que se muestran en la tabla 14.23.

TABLA 14.23 Pesos para el ejemplo 14.8

Capa	*Peso*	*Capa*	*Peso*
−7	1.4956	1	0.0845
−6	0.4119	2	0.0991
−5	0.2755	3	0.0997
−4	0.2304	4	0.0797
−3	0.1411	5	0.0534
−2	0.1349	6	0.0450
−1	0.1041	7	0.0520
0	0.0722		

Mediante el uso de estos pesos y al ajustar un modelo cuadrático, se obtienen, para el czono, los resultados que se encuentran en la tabla 14.24. Es evidente que una ecuación cuadrática de regresión basada en mínimos cuadrados con factores de peso es muy adecuada para describir la variabilidad de la cantidad promedio de ozono como una función de la altitud.

La mayoría de las veces no existen observaciones repetidas, pero los datos se recaban en agrupaciones naturales las que pueden, *a priori,* sugerir varianzas diferentes para el error en cada grupo. Lo que en general se hace es suponer que la varianza del j-ésimo grupo es $c_j^2\sigma^2$, donde c_j es única para el j-ésimo grupo, pero σ^2 es común para todos los grupos. En general, los valores de las c_j no son conocidos, pero pueden estimarse primero al determinar la varianza residual para cada grupo, s_j^2 basada en

TABLA 14.24 Estimaciones por mínimos cuadrados con factores de peso y tabla ANOVA para el ejemplo 14.8

Variable	*Coeficiente de regresión estimado*	*Desviación estándar estimada*	*Valor t*
Constante	96.7590	0.6367	151.98
x	−0.5585	0.1266	−4.41
x^2	−0.9495	0.0238	−39.83
$R^2 = 0.9817$		$t_{0.975,\,58} = 2.00$	

ANOVA

Fuente	gl	SC	CM	*Valor F*
Regresión	2	4082.33	2041.17	1556.30
Efecto lineal	1	2001.11	2001.11	1525.78
Efecto cuadrático	1	2081.22	2081.22	1586.82
Error	58	76.07	1.31	
Total	60	4158.40	$f_{0.95,\,2,\,58} = 3.15$; $f_{0.95,\,1,\,58} = 4.00$	

los residuos de éstos. Los residuos se obtienen mediante el ajuste de un modelo lineal general empleando mínimos cuadrados ordinarios. Entonces, una estimación de c_j es s_j/s, donde s es la desviación estándar residual global basada en los mínimos cuadrados ordinarios y s_j es la desviación estándar residual para el j-ésimo grupo. Entonces el peso para el j-ésimo grupo es $w_j = 1/c_j^2 = s^2/s_j^2$.

14.9 Variables indicadoras

En casi todos los problemas que se han considerado hasta este momento, las variables de predicción han sido cuantitativas en el sentido en que toman valores de una escala numérica bien definida. Sin embargo, para muchas variables como la localización geográfica, el estado civil, las poblaciones urbanas o rurales o alguna otra, no es evidente tener una escala bien definida. Dado que estas variables cualitativas son factores importantes en muchas situaciones, a continuación se examinará una manera de cuantificar los niveles de una variable de predicción cualitativa para su empleo en el análisis de regresión. Se considerarán las que comúnmente se conocen como variables indicadores o mudas. A cada una de estas variables se le asignan los valores 0 y 1.

Como ilustración, considérese la tasa de crímenes para dos estados adyacentes, para los que los datos figuran en el ejercicio 14.16 que se encuentra al final de este capítulo. En particular, supóngase que se desea hacer una regresión de la tasa de crímenes sobre el porcentaje de la población urbana en un estado para aquellos que se encuentren sólo en las regiones 1 y 5. El modelo de regresión será una función de la variable cuantitativa x_1 (porcentaje de población urbana) y una variable de predicción cualitativa que representa las dos regiones de interés.

Dado que sólo se tienen dos regiones, es conveniente definir dos variables indicadoras x_2 y x_3 tales, que

$$x_2 = \begin{cases} 1 & \text{si un estado es encuentra en la región 1,} \\ 0 & \text{de otro modo,} \end{cases}$$

$$x_3 = \begin{cases} 1 & \text{si un estado se encuentra en la región 5,} \\ 0 & \text{de otro modo.} \end{cases}$$

Entonces, para obtener una sola ecuación de regresión para ambas regiones, se deberá ajustar el modelo

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon.$$

Pero si se hace esto, entonces la matriz $\mathbf{X'X}$ no tendría inversa. Una manera fácil de salir de este problema es eliminar una de las dos variables indicadoras y emplear solamente una, por ejemplo x_2, en donde al igual que antes,

$$x_2 = \begin{cases} 1 & \text{si un estado se encuentra en la región 1,} \\ 0 & \text{de otro modo,} \end{cases}$$

En otra palabras, para cualquier estado que se encuentre en la región 1 ($x_2 = 1$) o si se encuentra en la región 5 ($x_2 = 0$). En general, si una variable cualitativa tiene m niveles, puede representarse por medio de m - 1 variables indicadoras, asignando a cada una los valores de 0 y 1.

Considérese de nuevo el problema de la tasa de crímenes para las regiones 1 y 5. Existen varias maneras de abordar el desarrollo de un modelo de regresión. Se puede reunir la información proveniente de ambas regiones y entonces ajustar el modelo lineal simple

$$Y = \beta_0 + \beta_1 x_1 + \varepsilon,$$

ignorando las diferencias regionales pueden obtenerse ecuaciones de regresión separadas para las regiones, cada una con diferentes estimaciones para los coeficientes de regresión. La elección entre estas dos opciones debe hacerse con mucho cuidado. En realidad debe decidirse si cada una de estas dos regiones es distinta con respecto a la tasa de crímenes, o si existe alguna relación en común. Si lo primero es cierto y se ajusta el modelo dado con anterioridad, entonces es probable que la tasa de crímenes en una región se encuentre sobreestimada mientras que para la otra ocurre lo contrario. Si existe una relación en común, entonces no es necesario tener dos ecuaciones de regresión separadas.

Al comparar los resultados que se obtienen con base en las ecuaciones de regresión separadas y la única para las regiones 1 y 5 mediante el empleo del porcentaje de población urbana como la única variable de predicción, se obtienen los datos que se encuentran en la tabla 14.25.

La comparación revela que las estimaciones para cada una de las pendientes son, en esencia, las mismas, pero las estimaciones de las intersecciones son significativamente diferentes. Nótese también que la ecuación de regresión única exhibe las propiedades menos deseables. De hecho, con esta última ecuación las tasas para los estados que se encuentran en las región 1 se sobreestiman, mientras que para los estados que se encuentran en la región 5 se subestiman con una sola excepción. Por lo tanto, en forma aparente existen diferencias regionales para la respuesta y no deben ignorarse.

Para incorporar las diferencias regionales dentro del modelo, sólo se utilizará la variable indicadora x_2 definida con anterioridad; el modelo se convierte en

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon. \tag{14.32}$$

Para interpretar los coeficientes de regresión, considérense los estados de la región 5. Dado que para éstos $x_2 = 0$, se supone una curva de regresión dada por

$$E(Y) = \beta_0 + \beta_1 x_1,$$

que es la ecuación de una línea recta con pendiente β_1 e intersección β_0. Para los estados que se encuentran en la región 1, $x_2 = 1$, y la respuesta media toma la forma

$$\begin{aligned} E(Y) &= \beta_0 + \beta_1 x_1 + \beta_2 \\ &= (\beta_0 + \beta_2) + \beta_1 x_1, \end{aligned}$$

TABLA 14.25 Modelos de regresión combinado y separado para el ejemplo de la tasa de crímenes

Modelo de regresión estimado

Variable	*Coeficiente de regresión estimado*	*Desviación estándar estimada*	*Valor t*
Constante	7.0350	4.2300	1.66
x_1	−0.0094	0.0673	−0.14
$n = 12$	$r^2 = 0.002$	$t_{0.975,\,10} = 2.228$	

Modelo de regresión para la región 1

Variable	*Coeficiente de regresión estimado*	*Desviación estándar estimada*	*Valor t*
Constante	0.4170	0.8020	0.52
x_1	0.0404	0.0118	3.41
$n = 6$	$r^2 = 0.745$	$t_{0.975,\,4} = 2.776$	

Modelo de regresión para la región 5

Variable	*Coeficiente de regresión estimado*	*Desviación estándar estimada*	*Valor t*
Constante	7.4400	3.9500	1.88
x_1	0.0439	0.0686	0.64
$n = 6$	$r^2 = 0.093$	$t_{0.975,\,4} = 2.776$	

la que también es la ecuación de una línea recta con la misma pendiente, β_1, pero con una intersección diferente $\beta_0 + \beta_2$. Entonces el modelo dado por (14.32) proporciona la respuesta promedio como una función lineal de x_1 con la misma pendiente para ambas regiones, pero con diferentes intersecciones. El parámetro β_2 representa el efecto diferencial que existe entre las intersecciones de las dos regiones. Para ajustar el modelo (14.32) el vector Y y la matriz X son

$$\mathbf{Y} = \begin{bmatrix} 4.2 \\ 2.4 \\ 3.1 \\ 3.2 \\ 3.9 \\ 1.4 \\ 10.2 \\ 11.7 \\ 10.6 \\ 11.9 \\ 9.0 \\ 6.0 \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & 77.6 & 1 \\ 1 & 50.8 & 1 \\ 1 & 84.6 & 1 \\ 1 & 56.4 & 1 \\ 1 & 87.1 & 1 \\ 1 & 32.2 & 1 \\ 1 & 80.5 & 0 \\ 1 & 60.3 & 0 \\ 1 & 45.0 & 0 \\ 1 & 47.6 & 0 \\ 1 & 63.1 & 0 \\ 1 & 39.0 & 0 \end{bmatrix}$$

Los resultados de la regresión se muestran en la tabla 14.26

TABLA 14.26 Análisis de regresión para el ejemplo de la tasa de crímenes

Variable	Coeficiente de regresión estimado	Desviación estándar estimada	Valor t
Constante	7.5800	1.6400	4.62
x_1	0.0416	0.0269	1.54
x_2	−7.2340	0.9520	−7.60
$n = 12$	$R^2 = 0.865$	$t_{0.975,\,9} = 2.262$	

Con base en estos resultados, las diferencias regionales son estadísticamente significativas. De esta forma, la última ecuación de regresión es superior con respecto al modelo único en el cual no se consideraban las diferencias regionales. En particular, las dos regiones tienen la misma estimación para la pendiente (0.0416), pero las intersecciones son iguales a 7.58 para la región 5 y 7.58 − 7.23 = 0.35 para la región 1. En general puede suponerse que la pendiente es la misma y, por lo tanto, es mejor emplear un modelo con una variable indicadora que un modelo único. Además, también es mejor un modelo con una variable indicadora que emplear dos modelos de regresión separados debido a que para el primero se tiene un mayor número de grados de libertad disponible para el error que para el segundo. De acuerdo con lo anterior, β_0 y β_2 son las estimaciones con la mejor precisión como es el caso para este ejemplo.

¿Qué ocurre si la pendiente no es la misma? Esta situación puede manejarse mediante la introducción en el modelo de un término de interacción para la variable cuantitativa x_1 y para la variable indicadora x_2. El modelo propuesto se convierte en

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon. \tag{14.33}$$

Para los estados que se encuentran en la región 5, $x_2 = 0$. Entonces, $x_1 x_2 = 0$, y la respuesta promedio para esta región es

$$E(Y) = \beta_0 + \beta_1 x_1.$$

Para los estados que se encuentran en la región 1, $x_2 = 1$, y $x_1 x_2 = x_1$, la respuesta media para esta región es

$$\begin{aligned} E(Y) &= \beta_0 + \beta_1 x_1 + \beta_2 + \beta_{12} x_1 \\ &= (\beta_0 + \beta_2) + (\beta_1 + \beta_{12}) x_1. \end{aligned}$$

Nótese que el coeficiente de regresión de x_2 es el efecto diferencial que existe entre las intersecciones de las dos regiones y el coeficiente de regresión del producto cruzado $x_1 x_2$ es el efecto diferencial entre las pendientes de las dos regiones. Por lo tanto, suponiendo que existe una interacción estadísticamente apreciable entre x_1 y x_2, pueden obtenerse las ecuaciones estimadas de regresión para cada región mediante el empleo del modelo dado por (14.33).

Para finalizar, se examinará el problema en el cual una variable cualitativa tiene más de dos niveles. Este caso requiere del uso de más de una variable indicadora en

el modelo de regresión. Como ilustración, se continuará con el problema de la tasa de crímenes al llevar a cabo una comparación entre los estados de las regiones 1, 5 y 7. Dado que se han identificado tres niveles de una variable cualitativa, se definirán dos variables indicadoras de la siguiente manera:

$$x_2 = \begin{cases} 1 & \text{si un estado se encuentra en la región 1,} \\ 0 & \text{de otro modo,} \end{cases}$$

$$x_3 = \begin{cases} 1 & \text{si un estado se encuentra en la región 5,} \\ 0 & \text{de otro modo.} \end{cases}$$

Este arreglo proporciona el mismo número de combinaciones posible de los valores de x_2 y x_3 de acuerdo con el número de niveles de la variable cualitativa. Estos son $x_2 = 1, x_3 = 0$; $x_2 = 0, x_3 = 1$; y $x_2 = x_3 = 0$. Representan los estados en las regiones 1, 5 y 7 respectivamente.

Si se supone que las pendientes son iguales para las tres regiones, el modelo es

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon.$$

Para los estados que se encuentran en la región 7, $x_2 = 0$ y $x_3 = 0$, de tal manera que la respuesta se reduce a

$$E(Y) = \beta_0 + \beta_1 x_1,$$

que es la ecuación de una línea recta con pendiente β_1 e intersección β_0. Para los estados que se encuentran en la región 5, $x_2 = 0$ y $x_3 = 1$. De acuerdo con lo anterior, la curva de regresión es

$$\begin{aligned} E(Y) &= \beta_0 + \beta_1 x_1 + \beta_3 \\ &= (\beta_0 + \beta_3) + \beta_1 x_1, \end{aligned}$$

donde β_3 representa el cambio en la intersección de la región 5 con respecto al de la región 7. De manera similar, cuando $x_2 = 1$ y $x_3 = 0$, la respuesta media es

$$\begin{aligned} E(Y) &= \beta_0 + \beta_1 x_1 + \beta_2 \\ &= (\beta_0 + \beta_2) + \beta_1 x_1, \end{aligned}$$

donde ahora β_2 es el cambio en la intersección de la región 1 con respecto al de la región 7. Se deduce que tanto β_2 como β_3 representan los efectos diferenciales para las intersecciones de las regiones 1 y 5, respectivamente, en relación con la intersección de la región 7.

El caso para el cual no es posible suponer que las pendientes son iguales, en este momento debe ser ya evidente, es decir, si se asume un modelo de la forma

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \varepsilon, \qquad (14.34)$$

donde β_{12} y β_{13} son los coeficientes de regresión para las interacciones que

comprenden a la variable cuantitativa x_1 y a cada una de las dos variables indicadoras x_2 y x_3.

Ejemplo 14.9 Se seleccionan al azar cinco casas recientemente vendidas para tres distintas zonas residenciales (A, B y C) en cierta ciudad, y el precio de venta Y se compara con el valor catastral de la propiedad x_1 determinado por la oficina estatal local correspondiente. Los datos se encuentran en la tabla 14.27 donde el precio de venta y el valor catastral de la propiedad se dan en miles de dólares. Mediante el empleo de variables indicadoras, ajústese una ecuación de regresión lineal y determínese si las pendientes para las tres zonas residenciales son las mismas.

Dado que se tienen tres zonas residenciales, se definen dos variables indicadoras x_2 y x_3 tales, que

$$x_2 = \begin{cases} 1 & \text{si una casa se encuentra en la zona B,} \\ 0 & \text{de otro modo,} \end{cases}$$

$$x_3 = \begin{cases} 1 & \text{si una casa se encuentra en la zona C,} \\ 0 & \text{de otro modo.} \end{cases}$$

Para el modelo (14.34) el vector Y y la matriz X son iguales a

$$\mathbf{Y} = \begin{bmatrix} 42.5 \\ 36.8 \\ 42.6 \\ 41.2 \\ 48.6 \\ 75.2 \\ 83.4 \\ 83.3 \\ 116.8 \\ 114.3 \\ 122.8 \\ 125.6 \\ 132.5 \\ 127.4 \\ 147.8 \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & 33.1 & 0 & 0 & 0 & 0 \\ 1 & 42.0 & 0 & 0 & 0 & 0 \\ 1 & 47.8 & 0 & 0 & 0 & 0 \\ 1 & 53.4 & 0 & 0 & 0 & 0 \\ 1 & 59.6 & 0 & 0 & 0 & 0 \\ 1 & 63.9 & 1 & 0 & 63.9 & 0 \\ 1 & 68.4 & 1 & 0 & 68.4 & 0 \\ 1 & 72.3 & 1 & 0 & 72.3 & 0 \\ 1 & 77.8 & 1 & 0 & 77.8 & 0 \\ 1 & 80.8 & 1 & 0 & 80.8 & 0 \\ 1 & 96.5 & 0 & 1 & 0 & 96.5 \\ 1 & 101.8 & 0 & 1 & 0 & 101.8 \\ 1 & 106.2 & 0 & 1 & 0 & 106.2 \\ 1 & 112.6 & 0 & 1 & 0 & 112.6 \\ 1 & 120.5 & 0 & 1 & 0 & 120.5 \end{bmatrix}$$

TABLA 14.27 Datos de la muestra para el ejemplo 14.9

Zona A		*Zona* B		*Zona* C	
x	*Y*	*x*	*Y*	*x*	*Y*
33.1	42.5	63.9	75.2	96.5	122.8
42.0	36.8	68.4	83.4	101.8	125.6
47.8	42.6	72.3	83.3	106.2	132.5
53.4	41.2	77.8	116.8	112.6	127.4
59.6	48.6	80.8	114.3	120.5	147.8

El listado de computadora que produce Minitab se encuentra en la figura 14.11, donde C2-C6 se refieren a x_1, x_2, x_3, x_1x_2, y x_1x_3, respectivamente.

Nótese que la hipótesis nula

$$H_0: \beta_{12} = 0$$

puede rechazarse, pero la hipótesis

$$H_0: \beta_{13} = 0$$

no; por lo tanto, existe una razón para creer que las pendientes para las zonas residenciales A y B no son las mismas. Del listado se determina que las ecuaciones esti-

---

```
LA ECUACION DE REGRESION ES
Y =   31.4   + 0.232 X1  -  129.X2
   +  1.89 X3    + 2.41 X4+ 0.679 X5
```

	COLUMNA	COEFICIENTE	DEV. EST. DEL COEF.	COCIENTE-T = COEF/D.E.
	—	31.37	14.66	2.14
X1	C2	0.2325	0.3050	0.76
X2	C3	-128.81	36.29	-3.55
X3	C4	1.89	38.82	0.05
X4	C5	2.4112	0.5481	4.40
X5	C6	0.6786	0.4518	1.50

LA DEV. EST. DE Y CON RESPECTO A LA RECTA DE REGRESION ES
S = 6.238
CON (15 - 6) = 9 GRADOS DE LIBERTAD

R-CUADRADA = 98.4 POR CIENTO

ANALISIS DE VARIANZA ;E

DEBIDO A	GL	SC	CM = SC/GL
REGRESION N	5	21577.96	4315.59
RESIDUO	9	350.26	38.92
TOTAL	14	21928.22	

ANALISIS DE VARIANZA ADICIONAL
SC EXPLICADA POR CADA VARIABLE QUE ENTRA EN EL ORDEN DADO

DEBIDO A	GL	SC
REGRESION	5	21577.96
C2	1	19892.61
C3	1	698.16
C4	1	232.89
C5	1	666.52
C6	1	87.79

FIGURA 14.11 Listado de computadora para el ejemplo 14.9

madas de regresión para cada zona residencial son las siguientes:

$$\text{Zona A:} \quad \hat{y} = 31.37 + 0.2325x_1, \quad (x_2 = x_3 = 0)$$

$$\text{Zona B:} \quad \hat{y} = -97.44 + 2.6437x_1, \quad (x_2 = 1, x_3 = 0)$$

$$\text{Zona C:} \quad \hat{y} = 33.26 + 0.9111x_1. \quad (x_2 = 0, x_3 = 1)$$

Referencias

1. S. Chatterjee and B. Price, *Regression analysis by example*, Wiley, New York, 1977.
2. C. Daniel and F. S. Wood, *Fitting equations to data*, Wiley, New York, 1971.
3. N. R. Draper and H. Smith, *Applied regression analysis*, 2nd ed., Wiley, New York, 1981.
4. F. A. Graybill, *Theory and application of the linear model*, Duxbury, North Scituate, Mass., 1976.
5. J. Neter, W. Wasserman, and M. H. Kutner, *Applied linear regression models*, Richard D. Irwin, Homewood, Ill., 1983.
6. *SAS users guide*, SAS Institute, Raleigh, N. C., 1982.

Ejercicios

14.1. De los siguientes modelos, ¿cuáles no son casos del modelo lineal general y por que?

a) $Y = \beta_0 + \beta_1 \exp(\beta_2 x_1) + \beta_3 x_2 + \varepsilon$

b) $Y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_1^2 x_2 + \varepsilon$

c) $Y = \beta_0 + \beta_1/x_1 + \beta_2 x_2^{1/2} + \varepsilon$

14.2. Una agencia de alquiler de automóviles obtiene la siguiente ecuación de regresión:

$$\hat{y} = 0.75 + 1.2x_1 + 0.15x_2$$

para predecir el costo promedio anual Y en miles de dólares como una función del número de automóviles alquilados x_1 y del número promedio de millas que cada automóvil recorre x_2 en miles de millas. Explíquese el significado de cada uno de los coeficientes estimados de la regresión.

14.3. Supóngase que la ecuación estimada de regresión que describe una respuesta media como una función de dos variables de predicción está dada por

$$\hat{y} = 15 + 6x_1 - 2x_2 - 1.5x_1x_2.$$

a) ¿Cúal es el efecto sobre la respuesta media por unidad de cambio en x_1 cuando x_2 = 2?

b) ¿Cuál es el efecto sobre la respuesta media por unidad de cambio en x_2 cuando x_1 = 1?

14.4. Mediante el empleo de los datos de Prater, ajústense todos los modelos lineales que incluyan sólo a x_1 y a x_3, e ilústrese el principio de la suma de cuadrados extra mediante el cálculo de lo siguiente:

a) Las tablas de análisis de varianza correspondientes.
b) $SCR(x_3 \mid x_1)$ y $SCR(x_1 \mid x_3)$.
c) Las pruebas F parciales apropiadas.

14.5. Una agencia desea estimar los gastos en alimentación de una familia con base en el ingreso y su tamaño. Los datos que se encuentran en la tabla 14.28 representan los gastos de alimentación por mes Y en miles de dólares, contra el ingreso mensual x_1 y el tamaño de la familia x_2 para 15 familias que se seleccionaron al azar en cierta localidad geográfica.

TABLA 14.28 Datos de la muestra para el ejercicio 14.5

Y	x_1	x_2
0.43	2.1	3
0.31	1.1	4
0.32	0.9	5
0.46	1.6	4
1.25	6.2	4
0.44	2.3	3
0.52	1.8	6
0.29	1.0	5
1.29	8.9	3
0.35	2.4	2
0.35	1.2	4
0.78	4.7	3
0.43	3.5	2
0.47	2.9	3
0.38	1.4	4

a) Ajústense todos los modelos lineales que abarcan a x_1 y/o x_2, e interprétense los coeficientes de regresión estimados.
b) Pruébese la hipótesis nula H_0: $\beta_1 = \beta_2 = 0$.
c) Calcúlese $SCR(x_2 \mid x_1)$ y $SCR(x_1 \mid x_2)$ y llévense a cabo las pruebas F parciales apropiadas.
d) Calcúlese e interprétese el coeficiente de correlación múltiple R^2.
e) Con base en los resultados anteriores, decídase cuál es la mejor ecuación para predecir el gasto de alimentación y empléese para estimar el gasto promedio mensual en alimentación para una familia de cuatro personas con un ingreso mensual de $2 500. Determínese un intervalo de confianza del 98% para esta cantidad.

14.6. Con respecto al ejercicio 14.5 hágase lo siguiente:

a) Para la regresión que comprende, tanto a x_1 como a x_2, efectúense las pruebas individuales t para los coeficientes de regresión β_1 y β_2. Úsese $\alpha = 0.05$.
b) Determínense intervalos de confianza de 95% para β_1 y β_2 y formúlense las conclusiones apropiadas.

14.7. Mediante el uso de los datos del ejercicio 14.5, constrúyase un modelo lineal general que abarque tanto a x_1 como a x_2 en forma matricial; identifíquense todas las matrices y obténganse las ecuaciones normales.

14.8. En muchas agencias gubernamentales y compañías privadas el problema de identificar aquellos factores que son importantes para predecir la aptitud para el trabajo de los aspirantes a obtener un ejemplo constituye un proceso continuo. El procedimiento usual es el de aplicar al solicitante un conjunto de pruebas apropiadas y tomar las decisión de contratarlo o no con base en los resultados de éstas. El asunto clave es conocer *a priori* qué pruebas pueden predecir la aptitud para el trabajo de una persona. Supóngase que el personal de una compañía muy grande ha desarrollado cuatro pruebas para una determinada clasificación con respecto al trabajo. Estas pruebas se aplicaron a 20 individuos que fueron contratados por la compañía. Después de un periodo de dos años, cada uno de estos empleados se clasifica de acuerdo con su aptitud para el trabajo. La puntuación para la aptitud hacia el trabajo Y y la correspondiente a cada una de las cuatro pruebas x_1, x_2, x_3, x_4 se dan en la tabla 14.29.

TABLA 14.29 Datos de la muestra para el ejercicio 14.8

Empleado	Y	x_1	x_2	x_3	x_4
1	94	122	121	96	89
2	71	108	115	98	78
3	82	120	115	95	90
4	76	118	117	93	95
5	111	113	102	109	109
6	64	112	96	90	88
7	109	109	129	102	108
8	104	112	119	106	105
9	80	115	101	95	88
10	73	111	95	95	84
11	127	119	118	107	110
12	88	112	110	100	87
13	99	120	89	105	97
14	80	117	108	99	100
15	99	109	125	108	95
16	116	116	122	116	102
17	100	104	83	100	102
18	96	110	101	103	103
19	126	117	120	113	108
20	58	120	77	80	74

a) Utilícese la rutina *PROC GLM* de *SAS* (o algún otro paquete comparable) para ajustar la regresión lineal de Y sobre x_1, x_2, x_3 y x_4.
b) Con base en el listado de la computadora que se obtiene en la parte *a*, prepárese una tabla de análisis de varianza mostrando todas las posibles pruebas F parciales.
c) Interprétense los coeficientes de regresión estimados y el coeficiente de correlación múltiple.

14.9. Empléense los datos del ejercicio 14.8 para hacer lo siguiente:

a) Obténganse todas las posibles ecuaciones de regresión, y para cada una calcúlese la

suma de los cuadrados de los errores, el cuadrado medio del error, el valor de C_p y el valor R^2 (véase el ejercicio 14.12).

b) Demuéstrese que la regresión por pasos y el procedimiento de eliminación hacia atrás proporcionan los mismos resultados para la mejor ecuación de predicción.

c) Con base en los resultados anteriores, dedúzcase la mejor ecuación de predicción y empléese para estimar la aptitud para el trabajo de un individuo que tiene las siguientes puntuaciones, en las pruebas: $x_1 = 105$, $x_2 = 110$, $x_3 = 99$, y $x_4 = 107$. Obténgase un intervalo de predicción del 95% para esta cantidad.

14.10. De manera reciente, se ha dirigido el interés hacia el desarrollo de métodos más rápidos y económicos para vigilar la concentración de sedimentos y contaminantes en los recursos acuíferos de cierta nación. Para los encargados de vigilar el medio ambiente, el interés principal recae en la necesidad de cuantificar los valores de concentración en el agua con base en datos de percepción remota. El uso de las técnicas de percepción remota para vigilar distintos parámetros que miden la calidad del agua parece ser prometedor. Un tipo de sistema de percepción remota es la variedad pasiva el cual depende, en forma única, de la radiación de sol como fuente de energía y mide el flujo total de radiación emitido por el sistema agua-atmósfera. Una componente muy grande de este flujo de radiación es el flujo de luz emitido por el agua, el cual, bajo condiciones normales, es una función de los constituyentes que se encuentran presentes en el agua. Para medir el espectro de esta radiación se encuentran disponibles un gran número de sistemas de rastreo multiespectral. Sin embargo, cada sistema tiene diferentes localizaciones de las bandas y anchos diferentes.

Se piensa que un cambio en la concentración de un contaminante causará un cambio en el valor del flujo de radiaciones, es decir, si se conocen los valores de la radiación para diferentes bandas espectrales, entonces es posibles predecir la concentración de un contaminante en una fuente de agua dada. El problema reside en el hecho de identificar, de entre todas las bandas, cuál es la que puede predecir la concentración del contaminante. En una tesis doctoral reciente, Whitlock* obtuvo datos reales de percepción remota proporcionados por un laboratorio, bajo condiciones controladas, que empleó cinco bandas y varios constituyentes, entre ellos el sedimento del feldespato. Los datos de la muestra se proporcionan en la tabla 14.30.

a) Empléense las cinco bandas como variables de predicción y las concentraciones de feldespato como la respuesta, para ajustar un modelo de regresión lineal.

b) Calcúlese la matriz de correlación para las cinco bandas de radiación. Interprétese el resultado.

c) Úsese la regresión por pasos y el procedimiento de eliminación hacia atrás, para determinar el mejor conjunto de variables de predicción. ¿Son los resultados idénticos?

d) Con base en los resultados anteriores, analícese cualquier aspecto que se considera evidente con respecto a este problema y que sirva para decidir por una ecuación de predicción adecuada.

14.11. En la sección 14.5 se mencionó que el valor de R^2 aumenta conforme se añaden más términos a la ecuación de regresión debido a que *SCE* siempre disminuye al sumar más términos y *STC* siempre permanece constante. Es por esta razón que se sugiere una medida alternativa que tome en cuenta el número de términos que figuran en el modelo.

* Charles H. Whitlock, tesis doctoral, Universidad Old Dominion, mayo, 1977.

TABLA 14.30 Datos de la muestra para el ejercicio 14.10

Concentración Y de feldespato	*Bandas de radiación*				
	x_1	x_2	x_3	x_4	x_5
17	0.297	0.310	0.290	0.220	0.156
17	0.360	0.390	0.369	0.297	0.205
35	0.075	0.058	0.047	0.034	0.023
69	0.114	0.100	0.081	0.058	0.042
69	0.229	0.213	0.198	0.142	0.102
173	0.315	0.304	0.267	0.202	0.147
173	0.477	0.518	0.496	0.395	0.285
17	0.072	0.063	0.047	0.036	0.024
17	0.099	0.092	0.074	0.056	0.038
73	0.420	0.452	0.425	0.332	0.235
17	0.189	0.178	0.153	0.107	0.076
35	0.369	0.391	0.364	0.286	0.200
69	0.142	0.124	0.105	0.077	0.056
35	0.094	0.087	0.072	0.049	0.032
35	0.171	0.161	0.145	0.094	0.068
52	0.378	0.420	0.380	0.281	0.200

Esta medida recibe el nombre de *coeficiente de correlación múltiple ajustado* y se define por

$$R_A^2 = 1 - \left(\frac{n-1}{n-p}\right)\frac{\text{SCE}}{\text{STC}}$$

donde p es el número de términos que contiene el modelo, incluyendo al término constante. Use la información que se encuentra en la tabla 14.10 para demostrar que el coeficiente de correlación múltiple ajustado para una regresión lineal que contiene a x_1, x_2, y x_3 es más pequeña que la de la ecuación que sólo contiene a x_2 y a x_3.

14.12. Empléese la ecuación de regresión estimada en el ejercicio 14.9, parte *a*, que incluye sólo a x_1, x_2, y x_4 para calcular los residuos estandarizados. Entonces, grafíquense contra los valores correspondientes de x_3. Explíquese el resultado.

14.13. Úsese la ecuación de reducción simple estimada del eje. 13.14; calcúlense y grafíquense los residuos estandarizados frente al producto anual bruto (PAB) X. ¿Qué se puede concluir? Ajústese una nueva ecuación de reducción, como lo sugiere la gráfica residual, para demostrar el riesgo de la extrapolación al estimar el pago de los impuestos porcentuales si el PAB es $ 250 000.

14.14. Empléese la ecuación estimada de regresión lineal simple obtenida en el ejercicio 13.13 para calcular y graficar los residuos estandarizados contra los años de experiencia. Obténgase una nueva ecuación de regresión, calcúlense los nuevos residuos estandarizados y de nuevo grafíquense contra x. ¿Qué se puede concluir?

14.15. Los datos que se encuentran en la tabla 14.31 representan la temperatura atmosférica promedio Y en enero para 50 estaciones climatológicas situadas en el estado de Virginia, donde cada estación se identifica por medio de su latitud x_1, longitud x_2 y altitud x_3.

TABLA 14.31 Datos de la muestra para el ejercicio 14.15*

Estación número	Y	x_1	x_2	x_3
1	37.9	37.35	79.52	975
2	28.7	38.52	78.43	3535
3	38.3	37.08	77.95	440
4	37.3	37.53	79.68	870
5	31.5	37.08	81.33	3300
6	35.0	37.38	80.08	1890
7	36.0	38.03	78.52	870
8	37.4	36.83	79.37	700
9	40.4	37.28	75.97	11
10	35.8	37.77	78.15	300
11	35.3	38.47	78.00	420
12	33.2	38.45	78.93	1400
13	39.3	36.58	79.38	410
14	41.3	36.90	76.20	25
15	34.7	38.45	77.67	300
16	38.0	37.33	78.38	450
17	34.2	36.93	80.30	2600
18	35.4	38.30	77.47	100
19	35.7	37.37	80.87	1524
20	39.7	36.68	76.78	80
21	40.5	37.30	77.30	40
22	31.6	38.00	79.83	2238
23	40.0	37.08	76.35	10
24	36.1	37.78	79.43	1060
25	34.1	39.12	77.72	500
26	36.1	38.03	78.00	420
27	33.9	38.67	78.38	1200
28	36.6	37.33	79.20	916
29	37.1	36.70	79.88	760
30	28.6	38.42	79.58	2910
31	29.3	39.07	77.88	1720
32	37.4	37.70	78.30	300
33	40.5	36.90	76.20	22
34	38.9	37.58	75.82	300
35	34.4	36.75	83.03	1510
36	35.3	38.50	77.32	12
37	37.5	37.50	77.33	164
38	36.4	37.32	79.97	1149
39	35.0	36.88	81.77	1735
40	34.0	38.15	79.03	1385
41	33.3	38.65	78.72	1000
42	38.6	37.65	76.57	25
43	37.5	37.75	77.05	50
44	36.2	37.85	75.48	9
45	32.1	38.95	77.45	291
46	35.6	38.85	77.03	10
47	39.3	37.30	76.70	70
48	33.7	39.20	78.17	760
49	34.4	38.88	78.52	887
50	34.4	36.93	81.08	2450

* *Fuente: Monthly normals of temperature, precipitation and heating and cooling degree days 1941–70*, No. 81, NOAA, U. S. Department of Commerce.

a) Ajústese un modelo de regresión de segundo orden completo y llévense a cabo los análisis apropiados sobre sus resultados.

b) Úsense los medios apropiados para evaluar si todos los términos que aparecen en la ecuación estimada de regresión deben retenerse. Si no es así, proporciónense argumentos suficientes para la elección de una ecuación de predicción adecuada.

14.16. Los datos de la tabla 14.32 representan la tasa de crímenes Y por cada 100 000 habitantes para los 48 estados de Estados Unidos y algunas variables de predicción potenciales como el porcentaje de población urbana x_1, el porcentaje de la población minoritaria x_2, la tasa de desempleo x_3, el porcentaje de la población que tiene cuatro o más años de educación x_4 y la región geográfica x_5.

a) Empléese un procedimiento de regresión por pasos para obtener el mejor conjunto de variables de predicción por incluir en un modelo lineal.

b) Para la mejor ecuación de predicción, calcúlense los residuos estandarizados y grafíquense contra las regiones. ¿La dispersión de estos residuos es esencialmente igual para todas las regiones? Si no es así obténganse los pesos para cada región mediante el empleo del procedimiento sugerido en la sección 14.8 y después utilícese el método de mínimos cuadrados con factores de peso para obtener las estimaciones de los coeficientes de regresión del mejor conjunto de variables de predicción. Compárense los resultados con los que se obtienen al emplear el método ordinario de mínimos cuadrados.

14.17. Emplénse los datos del ejercicio 14.16 para obtener la matriz de correlación para todas las variables potenciales de predicción y la respuesta. ¿Cuáles variables de predicción son las que tienen mayor correlación con la respuesta? ¿Está este resultado de acuerdo con la parte *a* del ejercicio 14.16? ¿Existen otras variables de predicción que se encuentren muy correlacionadas? Coméntese en términos del problema de la multicolinealidad.

14.18. Se cree que los salarios Y, en miles de dólares, para los profesores de cierta universidad por año académico están influenciados por tres variables: los años de experiencia en la enseñanza x_1; el rango académico x_2, y la disciplina x_3. Los datos que figuran en la tabla 14.33 provienen de una muestra aleatoria de 18 profesores de esta universidad. Los rangos académicos se identifican por un 1 para profesor asistente, 2 para profesor asociado y 3 para profesor titular. Las disciplinas se identifican mediante un 1 para ciencias, 2 para humanidades, 3 para artes y 4 para finanzas.

a) Defínanse las variables indicadoras para el rango y la disciplina. Entonces, ajústese un modelo lineal con el salario como la respuesta, los años de experiencia en la enseñanza, como la variable cuantitativa y las variables indicadoras representan el rango académico y la disciplina.

b) Interprétense los coeficientes de la regresión estimada.

c) Ajústese un modelo lineal que incluya todos los términos que contienen productos cruzados entre las variables indicadoras y x_1. Llévese a cabo un análisis completo sobre esta ecuación de regresión y obténganse las conclusiones adecuadas.

d) Para cada disciplina y rango académico, grafíquese la ecuación de regresión estimada obtenida en la parte *c* como una función de x_1.

TABLA 14.32 Datos de la muestra para el ejercicio 14.16

Estado	Y	x_1	x_2	x_3	x_4	x_5
1	14.2	58.4	25.8	7.4	10.2	6
2	9.5	79.6	9.2	9.8	15.7	8
3	8.8	50.0	18.4	6.6	9.1	6
4	11.5	90.9	12.0	8.2	16.8	8
5	6.3	78.5	4.7	5.6	19.4	7
6	4.2	77.4	6.6	7.1	18.3	1
7	6.0	72.2	15.2	8.9	15.5	2
8	10.2	80.5	14.9	9.0	13.7	5
9	11.7	60.3	26.5	6.9	12.3	5
10	5.5	54.1	1.8	6.3	13.5	7
11	9.9	83.0	14.7	6.5	13.7	3
12	7.4	64.9	7.6	5.7	11.0	3
13	2.3	57.2	1.6	4.0	12.8	4
14	6.6	66.1	5.6	4.0	14.6	4
15	10.1	52.3	7.5	4.6	10.0	6
16	15.5	66.1	30.2	7.0	11.5	6
17	2.4	50.8	0.7	8.9	13.6	1
18	8.0	76.6	21.1	6.8	18.6	2
19	3.1	84.6	4.3	9.5	16.8	1
20	9.3	73.8	12.5	8.2	12.6	3
21	2.7	66.4	2.0	5.9	13.2	4
22	14.3	44.5	36.4	7.4	11.5	6
23	9.6	70.1	11.2	6.2	11.8	4
24	5.4	53.4	4.8	6.2	14.2	7
25	3.9	61.5	3.8	5.0	12.8	4
26	15.8	80.9	8.3	9.0	13.1	8
27	3.2	56.4	0.7	6.4	15.3	1
28	5.6	88.9	12.8	9.4	14.9	2
29	8.8	69.8	9.8	7.8	15.3	8
30	10.7	88.9	14.6	9.1	16.0	2
31	10.6	45.0	23.1	6.2	11.8	5
32	0.9	44.3	3.3	5.5	12.2	4
33	7.8	75.3	10.1	7.8	11.5	3
34	8.6	68.0	11.3	5.0	11.7	6
35	4.9	67.1	3.0	9.5	15.4	7
36	5.6	71.5	9.4	7.9	11.9	2
37	3.9	87.1	3.7	8.6	14.9	1
38	11.9	47.6	31.2	5.0	10.4	5
39	2.0	44.6	6.1	3.6	11.4	4
40	10.1	58.7	15.9	6.0	10.5	6
41	13.3	79.7	13.1	5.7	13.7	6
42	3.5	80.4	2.5	5.3	17.5	7
43	1.4	32.2	0.8	8.0	15.6	1
44	9.0	63.1	19.5	5.6	16.4	5
45	4.3	72.6	5.1	8.8	16.1	7
46	6.0	39.0	3.9	7.5	9.2	5
47	2.8	65.9	3.9	4.5	12.7	3
48	5.4	60.5	3.1	3.6	14.5	7

Fuente: World almanac, 1979

TABLA 14.33 Datos de la muestra para el ejercicio 14.18

Y	x_1	x_2	x_3
25.7	10	1	2
18.8	4	1	1
18.6	5	1	3
21.8	13	2	3
26.3	4	1	4
29.4	24	3	3
28.6	7	2	2
34.5	12	3	4
24.3	11	2	1
21.2	6	1	3
28.8	6	1	4
24.7	4	1	2
32.4	12	3	2
33.4	20	3	1
27.4	11	2	1
29.8	6	2	4
31.4	11	2	4
27.7	8	2	3

CAPÍTULO QUINCE

Métodos no paramétricos

15.1 Introducción

Los procedimientos inferenciales que hasta este momento se han estudiado, con excepción de los límites de tolerancia independientes de distribución analizados en el capítulo 8, y de la estadística de Kolmogorov-Smirnov, presentada en el capítulo 10, necesitan de la especificación de una distribución para la población de interés. Por ejemplo, el procedimiento del análisis de varianza se hace posible al asumir que las observaciones provienen de distribuciones normales. De esta forma, la mayor parte de los procesos inferenciales que se han presentado representan estimaciones con respecto a los parámetros de la población de interés. Por esta razón, este tipo de inferencias reciben el nombre de *métodos paramétricos.*

Para muchos de los métodos inferenciales que se han examinado se ha hecho un intento por determinar su robustez, y en muchas ocasiones se ha encontrado que los métodos son razonablemente robustos con respecto a las distribuciones supuestas. No obstante, en general los métodos paramétricos son más sensibles a las suposiciones para muestras de tamaño pequeño y, para muchos de ellos, su aplicación se encuentra limitada a aquellas observaciones que tienen un carácter cuantitativo, es decir, se supone que lo que se observa es una cantidad numérica continua como el volumen de ventas semanal, la cantidad de cierta sustancia que se vacía en un recipiente, la resistencia de una muestra de metal y otros más.

Las observaciones de tipo cuantitativo se definen, en forma general, sobre un *intervalo* o sobre una escala de *proporciones.* Las mediciones que se definen en una escala de intervalo se pueden distinguir y ordenar en forma numérica, y sus diferencias son significativas. Un ejemplo clásico de una escala de intervalo es aquel que incluye la medición de la temperatura. Puede escogerse entre registrar la temperatura en grados Celsius (para los cuales el punto de congelación del agua es de 0°C) o en grados Fahrenheit (para los que el punto de congelación es de 32°F). De esta forma el origen de las escalas es diferente, pero el significado de la diferencia entre 10°C y 15°C es el mismo que tiene la diferencia entre 20°C y 25°C.

Si una medición reúne los requisitos de una escala de intervalo y además tiene un verdadero punto de origen, entonces la medición se define sobre una escala de pro-

porciones. Por ejemplo, las alturas, los pesos, las resistencias y otros se encuentran definidos sobre una escala de proporciones ya que tienen verdaderos puntos cero, sin importar la unidad de medición. Las escalas de intervalo y de proporción son verdaderamente cuantitativas. Para la mayor parte de los métodos paramétricos que se han presentado, como son la construcción de intervalos de confianza, la prueba de hipótesis estadísticas y el ajuste de ecuaciones son aplicables a todas aquellas observaciones que se encuentran definidas, por lo menos, sobre una escala de intervalo.

Sin embargo, en muchas situaciones lo que se observa tiene un carácter cualitativo (no cuantitativo) y, por lo tanto, no puede definirse sobre una escala de intervalo o de proporciones. Tales situaciones se encuentran con frecuencia en las ciencias sociales y en las encuestas de mercado. Por ejemplo, no es probable que al evaluar las preferencias del consumidor con respecto a una bebida, se adhieran a una escala numérica significativa, incluso si se le pidiese al consumidor su opinión con respecto a la bebida en una escala de cinco puntos, donde 1 y 5 pueden representar reacciones muy negativas o muy positivas, respectivamente, la escala es arbitraria. En otras palabras, los números no tienen ningún significado físico más allá que el de representar con un número más grande la respuesta más favorable para la bebida.

Las observaciones de este tipo pueden definirse sobre una escala *ordinal,* dado que la distancia entre dos puntos no es de consecuencia y sólo es importante el *orden* o *rango* de los números. En algunas ocasiones, las observaciones sólo pueden definirse sobre una escala *nominal* debido a que se emplea, ya sea un nombre (símbolo) o un número para clasificar una característica de interés, pero el principio de orden no es de consecuencia. Por ejemplo, las personas pueden clasificarse de acuerdo con su sexo. Pueden emplearse los símbolos M y H o utilizar números como 122 y 48 para denotar mujer u hombre. Las observaciones que se definen sobre escalas nominales son mediciones con pocas propiedades.

Se han desarrollado procedimientos inferenciales que no se encuentran sujetos a la forma de la distribución de la población de interés y no requieren, en forma necesaria, que las observaciones se definan por lo menos en una escala de intervalo. Estos procedimientos inferenciales se conocen como *métodos no paramétricos.* Dado que estos métodos no necesitan que se especifique la forma de la distribución de la población de interés, también se conocen como métodos *independientes de la distribución* (recuérdese, por ejemplo, los límites de tolerancia independientes de la distribución estudiados en el capítulo 8). En un sentido relativo, los métodos paramétricos requieren de pocas suposiciones y, la mayor parte de las veces, son más fáciles de aplicar que los procedimientos paramétricos que se han presentado en los capítulos anteriores; además, los métodos no paramétricos pueden aplicarse en aquellas situaciones para las que las observaciones se definen, por lo menos, en una escala de intervalo y, en ocasiones, sobre escalas nominales. Pero si las observaciones se definen por lo menos en una escala de intervalo y la distribución de la población de interés es normal, los métodos no paramétricos son menos eficientes comparados con los procedimientos paramétricos que se basan en la suposición de normalidad.

Se han desarrollado muchos métodos paramétricos en los que se han incluido procedimientos de análisis de varianza y de regresión. Las referencias citadas al final del capítulo proporcionan un panorama completo de todos los métodos no para-

métricos. El propósito de este capítulo radica sólo en introducir los conceptos básicos y presentar algunos métodos que son, en forma especial, útiles. Estos procedimientos no paramétricos son comparables con los métodos paramétricos para la prueba de hipótesis con respecto a las medias de dos distribuciones normales independientes (sección 9.6.2), la prueba de hipótesis con respecto a las medias para observaciones igualadas (sección 9.6.4), experimentos unifactoriales en diseños y en bloque completamente aleatorios (sección 12.4 y 12.5) y correlación lineal (sección 13.8).

15.2 Pruebas no paramétricas para comparar dos poblaciones con base en muestras aleatorias independientes

En la sección 9.6.2 se consideró el problema de comparar las medias de dos distribuciones cuando se supone que son normales. En esta sección se analizarán dos procedimientos no paramétricos para comparar las distribuciones de dos poblaciones: la prueba *U* de Mann-Whitney y la prueba de tendencias de Wald-Wolfowitz. La única suposición necesaria para su aplicación es que las distribuciones de interés sean continuas. De acuerdo con lo anterior, se supondrá que $X_1, X_2, \ldots, X_{n_1}$ y $Y_1, Y_2, \ldots, Y_{n_2}$ son muestras aleatorias independientes de dos poblaciones con distribuciones continuas.

15.2.1 Prueba de Mann-Whitney

Dadas muestras aleatorias independientes de dos poblaciones, considérese la prueba de la hipótesis nula de que las poblaciones tienen la misma distribución. La hipótesis puede establecerse como

$$H_0: f_1(x) \equiv f_2(y), \tag{15.1}$$

donde $f_1(x)$ y $f_2(y)$ son las correspondientes funciones de densidad de probabilidad. La hipótesis alternativa puede ser uni o bilateral. La hipótesis alternativa bilateral establece en forma sencilla que las distribuciones no son las mismas. Pero la hipótesis alternativa sólo implica un desplazamiento en la tendencia central de una distribución con respecto a la otra y no sugiere una diferencia en la forma o en la dispersión. En otras palabras, al igual que para el procedimiento *t* de Student, se supone que las distribuciones tienen la misma forma y dispersión.

Un procedimiento popular no paramétrico para probar la hipótesis nula dada por (15.1) es la *prueba U de Mann-Whitney.** Esta prueba es el equivalente no paramétrico de la prueba *t* de student para dos muestras estudiada en la sección 9.6.2. La prueba de Mann-Whitney se basa en una combinación de las n_1 y n_2 observaciones para formar un solo conjunto de $n_1 + n_2$ observaciones arregladas en orden creciente de magnitud. Entonces se asigna un *rango* a cada observación en la secuencia ordenada que comienza con un rango 1 y termina con un rango $n_1 + n_2$. Si las muestras aleatorias provienen de poblaciones que tienen la misma distribución, se espera que los rangos se encuentren lo suficientemente dispersos cuando se observa

* Este procedimiento es, en forma esencial, igual a la prueba de Wilcoxon de la suma del rango.

en qué muestra se encuentran las observaciones. De otra forma, debe esperarse que los rangos de las observaciones en cada muestra se encuentren muy agrupados en los extremos. En esencia, la estadística de Mann-Whitney determina cuándo un agregado de rangos observados es suficiente para concluir que las dos muestras aleatorias provienen de poblaciones cuyas distribuciones difieren en la tendencia central.

Para implementar el procedimiento se obtiene la suma de los rangos asociados con las observaciones de una de las dos muestras, por ejemplo la muestra 1, la cual se escoge en forma arbitraria. Denótese esta suma por R_1. Entonces la estadística U de Mann-Whitney está dada por

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1. \tag{15.2}$$

La estadística U es una función de la variable aleatoria R_1 y de los tamaños de las muestras n_1 y n_2. Si H_0 es cierta, la ocurrencia de cualquier orden particular para las observaciones en el conjunto combinado es equiprobable. Por lo tanto, bajo H_0, R_1 es la suma de n_1 enteros positivos seleccionados en forma aleatoria de entre los primeros $n_1 + n_2$. De acuerdo con lo anterior, puede determinarse que

$$E(R_1) = n_1(n_1 + n_2 + 1)/2, \tag{15.3}$$

$$Var(R_1) = n_1 n_2(n_1 + n_2 + 1)/12. \tag{15.4}$$

De (15.2) sigue que

$$E(U) = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - E(R_1) = n_1 n_2/2, \tag{15.5}$$

y

$$Var(U) = Var(R_1) = n_1 n_2(n_1 + n_2 + 1)/12. \tag{15.6}$$

Se ha determinado y tabulado la distribución exacta de U. Se invita al lector a que consulte [1] y [2] para conocer los detalles. Para una hipótesis alternativa bilateral, es probable que se rechace H_0 si se obtiene un valor muy grande o muy pequeño de U. Lo anterior ocurrirá cuando el valor de R_1 es muy grande o muy pequeño, respectivamente. Sin embargo, cuando tanto n_1 y n_2 son mayores de 10, la distribución de U se encuentra, en forma adecuada, aproximada por una distribución normal con media y varianza dadas por (15.5) y (15.6), respectivamente, es decir, bajo H_0 la variable aleatoria

$$Z = \frac{U - E(U)}{\sqrt{Var(U)}}$$

es aproximadamente $N(0, 1)$ para valores grandes de n_1 y n_2.

Debe notarse que a pesar de que no pueden ocurrir empates en la práctica desde un punto de vista teórico, esto ocurre en muchas ocasiones. Cuando ocurre un empate en la secuencia ordenada, se sugiere asignar el promedio de los rangos a las observaciones para las cuales ocurre el empate. Por ejemplo, supóngase que las obser-

vaciones octava y novena en la secuencia ordenada son las mismas. Entonces a cada una de estas observaciones se les asigna un rango de 8.5.

Ejemplo 15.1 **Se sospecha que una compañía lleva a cabo una política de discriminación, con respecto al sexo, en los salarios de sus empleados. Se seleccionaron 12 empleados masculinos y 12 femeninos de entre los que tienen responsabilidades y experiencia similares en el trabajo; sus salarios anuales en miles de dólares son los siguientes:**

Mujeres	22.5	19.8	20.6	24.7	23.2	19.2	18.7	20.9	21.6	23.5	20.7	21.6
Hombres	21.9	21.6	22.4	24.0	24.1	23.4	21.2	23.9	20.5	24.5	22.3	23.6

¿Existe alguna razón para creer que estas muestras aleatorias provienen de poblaciones con diferentes distribuciones? Úsese $\alpha = 0.05$.

Se combinan los salarios de las dos muestras para formar un solo conjunto de 24 salarios anuales. Entonces se ordenan los salarios y se les asigna un rango de la siguiente manera:

Sexo	*M*	*M*	*M*	*H*	*M*	*M*	*M*	*H*	*H*	*M*	*M*	*H*
Rango del salario	18.7 1	19.2 2	19.8 3	20.5 4	20.6 5	20.7 6	20.9 7	21.2 8	21.6 10	21.6 10	21.6 10	21.9 12

Sexo	*H*	*H*	*F*	*M*	*H*	*M*	*H*	*H*	*H*	*H*	*H*	*M*
Rango del salario	22.3 13	22.4 14	22.5 15	23.2 16	23.4 17	23.5 18	23.6 19	23.9 20	24.0 21	24.1 22	24.5 23	24.7 24

Para obtener la suma de los rangos se seleccionará la muestra de mujeres. De esta forma la suma de los rangos es

$$1 + 2 + 3 + 5 + 6 + 7 + 10 + 10 + 15 + 16 + 18 + 24 = 117,$$

y el valor de la estadística U de Mann-Whitney es

$$u = (12)(12) + \frac{12(13)}{2} - 117 = 105.$$

Dado que $E(U) = (12)(12)/2 = 72$ y $Var(U) = (12)(12)(25)/12 = 300$, mediante el empleo de la aproximación normal,

$$z = (105 - 72)/\sqrt{300} = 1.91$$

es un valor de una variable aleatoria normal estándar. Para $\alpha = 0.05$, los valores críticos son ± 1.96. Por lo tanto, no puede rechazarse la hipótesis nula de que las muestras aleatorias provienen de poblaciones con distribuciones idénticas.

15.2.2 Prueba de tendencias de Wald-Wolfowitz

Otro método no paramétrico que compara las distribuciones de dos poblaciones con base en muestras aleatorias independientes es la ***prueba de tendencias de Wald-Wolfowitz***. Para esta prueba, la hipótesis nula es que las dos muestras aleatorias provienen de poblaciones que tienen distribuciones idénticas, pero a diferenca de la prueba U de Mann-Whitney, no sugiere una diferencia sólo en la tendencia central; es decir, la hipótesis alternativa en la prueba de Wald-Wolfowitz es mucho más amplia. Ésta establece simplemente que las distribuciones difieren en algún aspecto como en la tendencia central, en la dispersión o la asimetría.

Al igual que en la prueba de Mann-Whitney, las observaciones en las dos muestras aleatorias se combinan y ordenan de acuerdo con sus magnitudes. Pero en lugar de considerar los rangos, el procedimiento de Mann-Wolfowitz busca el número de tendencias en la secuencia ordenada.

Definición 15.1 Se define una tendencia de longitud j como una secuencia de j observaciones, todas pertenecientes al mismo grupo, que se encuentran, ya sea precedidas o seguidas por observaciones que pertenecen a un grupo diferente.

Como ilustración, recuérdese la secuencia ordenada del ejemplo 15.1. La secuencia ordenada de acuerdo con el sexo de los empleados es la siguiente:

F F F M F F F M M F F M
M M F F M F M M M M M F

Para el sexo del empleado, la secuencia ordenada exhibe tendencias de M y H. En particular, la secuencia comienza con una tendencia de longitud tres, seguida por una tendencia de longitud uno, seguida por otra de longitud tres, y así consecutivamente. El número total de tendencias en esta secuencia es de 11.

Si la hipótesis nula de que las distribuciones son idénticas es cierta, las observaciones de las dos muestras en la secuencia ordenada deben encontrarse bien mezcladas, produciendo de esta forma un número grande de tendencias. Pero si las distribuciones de interés difieren en algún aspecto, es probable que la secuencia ordenada contenga tendencias de corta longitud obteniéndose de esta forma un número total de tendencias pequeño.

Sea R el número total de tendencias observadas en una secuencia ordenada de $n_1 + n_2$ observaciones, donde n_1 y n_2 son los respectivos tamaños de las muestras. Los posibles valores de R son $2, 3, \ldots, (n_1 + n_2)$. Puede demostrarse que la función de probabilidad de R está dada por

$$p(r) = \begin{cases} \dfrac{2\dbinom{n_1 - 1}{r/2 - 1}\dbinom{n_2 - 1}{r/2 - 1}}{\dbinom{n_1 + n_2}{n_1}} & r \text{ par}, \\[2ex] \dfrac{\dbinom{n_1 - 1}{r/2 - 1/2}\dbinom{n_2 - 1}{r/2 - 3/2} + \dbinom{n_1 - 1}{r/2 - 3/2}\dbinom{n_2 - 1}{r/2 - 1/2}}{\dbinom{n_1 + n_2}{n_1}} & r \text{ impar}. \end{cases} \tag{15.7}$$

La media y la varianza de R son

$$E(R) = \frac{2n_1n_2}{n_1 + n_2} + 1, \tag{15.8}$$

$$Var(R) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}. \tag{15.9}$$

Para probar H_0 con una probabilidad α, para el error de tipo I, debe encontrarse un entero r_0 tal, que en la medida de lo posible

$$\sum_{r=2}^{r_0} p(r) = \alpha.$$

Se rechaza la hipótesis nula cuando el número observado de tendencias es menor o igual a r_0. Nótese que la región crítica es una región unilateral inferior dado que se rechaza H_0 cuando el número de tendencias es bastante pequeño.

La distribución acumulativa de R se encuentra tabulada en forma extensa; pero si tanto n_1 como n_2 son mayores que 10, la distribución de R se encuentra, en forma adecuada, aproximada por una distribución normal con media y varianza dadas por (15.8) y (15.9), respectivamente. Como ilustración, recuérdese el ejemplo 15.1. El número observado de tendencias es 11, y para $n_1 = n_2 = 12$ los valores de la media y la varianza de R son 13 y 5.7391, respectivamente. Entonces, mediante el empleo de la aproximación normal,

$$z = (11 - 13)/\sqrt{5.7391} = -0.83$$

es un valor de una variable aleatoria normal estándar. Para $\alpha = 0.05$, se observa que la hipótesis nula no puede ser rechazada.

En la aplicación de la prueba de tendencias de Wald-Wolfowitz surge un problema muy serio cuando ocurren empates entre las observaciones que se encuentran en grupos diferentes. Este problema se debe a que el número de tendencias depende de cómo se manejen los empates en la secuencia ordenada. El procesamiento que se sugiere en estos casos es el de ordenar las observaciones empatadas en forma tal, que sea lo menos favorable para el rechazo de H_0. Pero si se tienen muchos empates, la validez de la prueba de Wald-Wolfowitz es cuestionable.

Por causa de la naturaleza extensa de la hipótesis alternativa en la prueba de Wald-Wolfowitz, ésta y la prueba de Mann-Whitney no son comparables. Si un investigador desea comparar las tendencias centrales de las distribuciones de dos poblaciones y sólo se tienen observaciones ordinales disponibles, la estadística de Mann-Whitney es el procedimiento no paramétrico más poderoso para detectar diferencias entre las tendencias centrales. Si se va a hacer una comparación más amplia, la prueba de Wald-Wolfowitz es un procedimiento viable pero menos poderoso.

15.3 Pruebas no paramétricas para observaciones por pares

En la sección 9.6.4 se consideró la comparación entre las medias de dos tratamientos cuando las observaciones se encuentran igualadas con el propósito de eliminar los

efectos causados por factores externos. En esta sección se discutirán dos pruebas no paramétricas que son equivalentes al procedimiento t de Student de la sección 9.6.4. Éstas son la *prueba del signo* y la *prueba de rangos y signos de Wilcoxon.*

15.3.1 La prueba del signo

La *prueba del signo* se basa en los signos de las diferencias entre las observaciones por pares de dos variables aleatorias X y Y. Sean (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n) pares de n observaciones muestrales de las distribuciones de X y Y, donde se supone que éstas son continuas. En muchas ocasiones existe una relación natural entre X y Y, por lo que X y Y no necesitan ser independientes. Por ejemplo, X y Y pueden representar las respuestas de parejas de matrimonios.

Para cada par en el que X es mayor que Y se registra un signo (+) de otra forma se registra un signo (–). Dado que se supone que las distribuciones de X y Y son continuas, en forma teórica, no pueden ocurrir empates. Sea p la probabilidad de que X sea mayor que Y. Entonces si la hipótesis nula es que X y Y tienen la misma distribución, el valor de p debe ser igual a 0.5. Sin embargo, debe notarse que p puede ser igual a 0.5, aun cuando las distribuciones de X y Y no sean idénticas. Por lo tanto, y en esencia, la hipótesis nula para la prueba del signo es

$$H_0: \quad p = 0.5,$$

La cual puede probarse contra hipótesis alternativas, ya sean uni o bilaterales, lo cual depende de lo que el investigador desee. Nótese que si H_0 es cierta, debe esperarse que, en forma aproximada, la mitad de los n pares tengan signos +.

La estadística para la prueba del signo, denotada por S, es el número de signos + entre los n pares. Dado que bajo H_0 cada par constituye un ensayo independiente con una probabilidad para el signo + de 0.5, la estadística S tiene una distribución binomial con $p = 0.5$. De acuerdo con lo anterior, para n dado y $p = 0.5$, se emplea la distribución binomial para obtener las regiones críticas de tamaño α para el error de tipo I. Para valores grandes de n puede utilizarse la aproximación normal de la distribución binomial, estudiada en la sección 5.2.

Cuando ocurren empates al aplicar la prueba del signo, el procedimiento que se recomienda seguir es el de ignorarlos y emplear la prueba sólo para aquellos pares en los que no ocurren empates. Este procedimiento puede representar un problema si se tienen numerosos empates y el número original de pares es relativamente pequeño.

Ejemplo 15.2 Se seleccionaron al azar 10 parejas de recién casados, y se les preguntó por separado, tanto al marido como a la esposa, cuántos hijos deseaban tener. Se obtuvo la siguiente información.

Pareja	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Esposa X	3	2	1	0	0	1	2	2	2	0
Marido Y	2	3	2	2	0	2	1	3	1	2

Mediante el empleo de la prueba del signo, ¿existe alguna razón para creer que las

esposas desean menos hijos que sus esposos? Supóngase un tamaño máximo del error del tipo I de 0.05.

Considérese la prueba de la hipótesis nula

$$H_0: \ p = 0.5$$

contra la alternativa

$$H_1: \ p < 0.5.$$

Nótese que deberá rechazarse H_0 si el número de signos + es muy pequeño. Al restar las respuestas de cada esposo de la de su esposa, y notando que las respuestas de cinco de las parejas son las mismas, se obtiene el siguiente arreglo de signos + y –:

Pareja	*1*	*2*	*3*	*4*	*6*	*7*	*8*	*9*	*10*
Signo	+	–	–	–	–	+	–	+	–

Existen tres signos + de manera tal, que el valor de la estadística S es 3. Dado que bajo H_0, S es binomial con $n = 10$ y $p = 0.5$, el valor p, o la probabilidad de observar tres o menos signos +, se obtiene de la tabla A del apéndice y es

$$P(S \leqslant 3) = 0.2539.$$

Dado que 0.2539 es mayor que $\alpha = 0.05$ la hipótesis nula no puede rechazarse. Nótese que para este ejemplo el valor crítico de S debe ser igual a uno si el tamaño máximo del error de tipo I es de 0.05.

15.3.2 Prueba de rangos de signos de Wilcoxon

La prueba del signo considera sólo las diferencias en el signo entre cada par de observaciones e ignora sus magnitudes. Si las observaciones se definen sobre una escala ordinal, las magnitudes de las diferencias tienen poco valor. Pero si las observaciones son magnitudes físicas, la prueba del signo puede ignorar mucha información debido a que no se toman en cuenta las magnitudes de las diferencias. La *prueba de rangos y de signos de Wilcoxon* toma en cuenta tanto el signo como la magnitud de las diferencias entre cada par de observaciones. Por lo tanto, para tener un buen balance, éste es el mejor método no paramétrico por utilizar para observaciones en parejas.

Para implementar la prueba de Wilcoxon, se obtienen las diferencias para los n pares de observaciones. Entonces, se ordenan sin importar el signo y de acuerdo con este orden se les asigna un rango, es decir, la diferencia más pequeña recibe un rango uno y a la diferencia absoluta más grande se le asigna un rango igual a n. Entonces, el signo de cada diferencia se une al rango de ésta. Los empates entre las diferencias se manejan de la misma manera que en la prueba de Mann-Whitney, pero si una diferencia es igual a cero, el procedimiento que se sugiere es omitir el par y ajustar n.

La estadística de la prueba de Wilcoxon es la suma de los rangos positivos y se denota por T_+. Nótese que T_+ contiene no sólo información proporcionada por la estadística de la prueba del signo sino también información con respecto a la magni-

tud relativa de las diferencias. Si la hipótesis nula de que las observaciones en cada par provienen de distribuciones idénticas es cierta, la ocurrencia de cualquier secuencia, en particular de los rangos y signos, es equiprobable de entre las 2_n secuencias posibles de signos + y –. Bajo la hipótesis nula, se espera que T_+ tenga el mismo valor, aproximadamente, que la suma de las magnitudes de los rangos negativos. Por lo tanto, dependiendo de la naturaleza de la hipótesis alternativa, se rechaza H_0 cuando se observa un valor de T_+ suficientemente grande o pequeño.

Se ha determinado y tabulado la distribución exacta de T_+. Sin embargo, al igual que para algunas otras estadísticas, la distribución de muestre de T_+ se encuentra aproximada, en forma adecuada, por una distribución normal para $n > 10$, donde

$$E(T_+) = n(n + 1)/4, \tag{15.10}$$

$$Var(T_+) = n(n + 1)(2n + 1)/24. \tag{15.11}$$

En otras palabras, la variable aleatoria

$$Z = \frac{T_+ - E(T_+)}{\sqrt{Var(T_+)}}$$

es aproximadamente $N(0, 1)$ para valores grandes de n.

Ejemplo 15.3 De una clase de estadística se seleccionan al azar 11 estudiantes y se observan sus calificaciones en dos exámenes sucesivos. Para las calificaciones dadas en la tabla 15.1, utilícese la prueba de rangos y de signos de Wilcoxon para determinar si el segundo examen fue más difícil que el primero. Úsese $\alpha = 0.1$.

En la tabla se encuentran las diferencias (examen 1 – examen 2), rangos, y rangos con signos para los 11 estudiantes. Dado que se desea determinar si el segundo examen fue más difícil que el primero, la hipótesis alternativa es unilateral, y la región crítica se encuentra en el extremo superior de la distribución de muestreo de T_+ es decir, si el valor observado de la suma de los rangos positivos es grande, lo anterior

TABLA 15.1 Datos de la muestra para el ejemplo 15.3

Estudiante	*Prueba 1*	*Prueba 2*	*Diferencia*	*Rango*	*Rango con signo*
1	94	85	9	8	8
2	78	65	13	10	10
3	89	92	−3	4	−4
4	62	56	6	7	7
5	49	52	−3	4	−4
6	78	74	4	6	6
7	80	79	1	1	1
8	82	84	−2	2	−2
9	62	48	14	11	11
10	83	71	12	9	9
11	79	82	−3	4	−4

implicaría tener calificaciones bajas, en forma suficiente, para el examen 2, y debe rechazarse la hipótesis nula de no diferencia.

La suma de los rangos positivos es $8 + 10 + 7 + 6 + 1 + 11 + 9 = 52$. Para $n = 11$, los valores de la media y la varianza de T_+ son $E(T_+) = 33$ y $Var(T_+) = 126.5$. Entonces, mediante el empleo de la aproximación normal,

$$z = \frac{52 - 33}{\sqrt{126.5}} = 1.69.$$

Para $\alpha = 0.1$, $z_{0.9} = 1.28$, y por lo tanto se rechaza la hipótesis nula.

15.4 Prueba de Kruskal-Wallis para k muestras aleatorias independientes

Recuérdese el procedimiento paramétrico del análisis de varianza de la sección 12.4, en el que el interés radica en probar la hipótesis nula

$$H_0: \quad \mu_1 = \mu_2 = \cdots = \mu_k,$$

con base en k muestras aleatorias mutuamente independientes provenientes de poblaciones cuyas distribuciones se suponen como normales. Se han desarrollado métodos no paramétricos para, de manera esencial, el mismo propósito siempre que por lo menos se encuentren disponibles mediciones ordinales y las distribuciones de las poblaciones de interés sean continuas. Uno de estos métodos es el procedimiento de *Kruskal-Wallis,* el cual prueba las hipótesis nulas de que los efectos de los tratamientos son los mismos, o que las k muestras aleatorias provienen de poblaciones con distribuciones idénticas.

Sean las observaciones de las k muestras aleatorias las dadas en la tabla 15.2, donde n_j es el tamaño de la j-ésima muestra y $N = \sum_{j=1}^{k} n_j$ es el número total de observaciones para todas las muestras.

La hipótesis nula puede establecerse como

$$H_0: \quad f_1(y) \equiv f_2(y) \equiv \cdots \equiv f_k(y) \tag{15.12}$$

donde $f_1(y), f_2(y), \cdots, f_k(y)$ son las correspondientes funciones de densidad de probabilidad. La hipótesis alternativa puede ser general y establecer sólo que las k

TABLA 15.2 Observaciones de k muestras aleatorias para la prueba de Kruskal-Wallis

			Muestra		
1	2	$\cdots$	j	$\cdots$	k
Y_{11}	Y_{12}	$\cdots$	Y_{1j}	$\cdots$	Y_{1k}
Y_{21}	Y_{22}	$\cdots$	Y_{2j}	$\cdots$	Y_{2k}
$\vdots$	$\vdots$		$\vdots$		$\vdots$
$Y_{n_1 1}$	$Y_{n_2 2}$	$\cdots$	$Y_{n_j j}$	$\cdots$	$Y_{n_k k}$

distribuciones no son idénticas. Sin embargo, la prueba de Kruskal-Wallis es sensible a las diferencias en tendencia central y es muy útil cuando se sospecha que las distribuciones de interés difieren sólo en ese aspecto. De acuerdo con lo anterior, el procedimiento de Kruskal-Wallis se considera, en general, como una extensión de la prueba U, de Mann-Whitney.

Al igual que en la prueba de Mann-Whitney, el procedimiento de Kruskal-Wallis se basa en la combinación de todas las observaciones en las muestras aleatorias para formar un solo conjunto de N observaciones; entonces, éstas se arreglan en orden creciente de magnitud y se asigna un rango a cada observación comenzando con un rango 1 y terminando con un rango N. Cuando el rango de todas las observaciones está completo, se determina la suma de los rangos para cada muestra. Sea R_j la suma de los rangos de la j-ésima muestra. En esencia, la prueba de Kruskal-Wallis determina si la disparidad entre las R_j con respecto a los tamaños n_j de las muestras es suficiente para garantizar el rechazo de la hipótesis nula.

Bajo la suposición de que las k muestras provienen de poblaciones con distribuciones idénticas, la estadística de la prueba de Kruskal-Wallis es

$$H = \frac{12}{N(N+1)}\left[\sum_{j=1}^{k}\frac{R_j^2}{n_j}\right] - 3(N+1), \qquad (15.13)$$

la que para tamaños n_j relativamente grandes de las muestras se encuentra aproximada, en forma adecuada, por una distribución chi-cuadrada con $k-1$ grados de libertad. Para un tamaño específico del error de tipo I, la región crítica es la porción superior de la distribución chi-cuadrada. De acuerdo con lo anterior, se rechaza la hipótesis nula para valores grandes de la estadística de la prueba de Kruskal-Wallis. Debe notarse que la aproximación chi-cuadrada es, por lo general, satisfactoria, excepto cuando $k = 3$ y ninguno de los tamaños de las muestras n_j sea mayor que cinco.

El procedimiento que se recomienda para manejar los empates es igual al de la prueba de Mann-Whitney. Si el número de empates es grande, se ha propuesto un factor de corrección para la estadística de pruebas dada por (15.3); véanse cualesquiera de las referencias que se encuentran al final de este capítulo. A pesar de que esta correción siempre incrementa el valor de la estadística de prueba, en muchos casos este efecto es despreciable, aun si existen numerosos empates.

Ejemplo 15.4 Se tomaron muestras aleatorias independientes de casas recientemente vendidas en cuatro zonas residenciales de una gran ciudad. El problema era determinar si existían diferencias en las zonas con respecto al valor de la propiedad y el precio de venta. Los datos que figuran en la tabla 15.3 son los cocientes entre los precios de venta y el valor catastral de la propiedad. Para $\alpha = 0.05$, empléese la estadística de Kruskal-Wallis para probar si estas muestras provienen de poblaciones con distribuciones idénticas.

Los valores que se encuentran entre paréntesis en la tabla son los rangos de las observaciones después de haberlas combinado y ordenado. Nótese que $n_1 = n_4 = 5$, $n_2 = n_3 = 6$ y $N = 22$. Las sumas de los rangos de cada muestra

TABLA 15.3 Datos de la muestra para el ejemplo 15.4

Zona residencial			
1	2	3	4
1.19 (15)	1.08 (4.5)	0.98 (2)	1.12 (7.5)
1.05 (3)	1.23 (17.5)	1.19 (15)	1.14 (10)
1.14 (10)	1:26 (20)	1.08 (4.5)	1.31 (22)
1.25 (19)	1.10 (6)	0.93 (1)	1.12 (7.5)
1.29 (21)	1.18 (12.5)	1.23 (17.5)	1.19 (15)
	1.14 (10)	1.18 (12.5)	

son $R_1 = 68$, $R_2 = 70.5$, $R_3 = 52.5$, y $R_4 = 62$. Entonces el valor de la estadística de Kruskal-Wallis es

$$h = \frac{12}{(22)(23)}\left[\frac{(68)^2}{5} + \frac{(70.5)^2}{6} + \frac{(52.5)^2}{6} + \frac{(62)^2}{5}\right] - 3(23) = 1.70.$$

De la tabla E del apéndice, para $\alpha = 0.05$ y $k - 1 = 3$ grados de libertad, el valor crítico es 7.82. Dado que $h = 1.70 < 7.82$, no puede rechazarse la hipótesis nula. Por lo tanto, no existe alguna razón para creer que existen diferencias entre las zonas cuando se comparan el precio de venta y el valor real de la propiedad.

15.5 Prueba de Friedman para k muestras igualadas

La prueba de rango de signos de Wilcoxon se considera como el equivalente no paramétrico del método t de Student para observaciones por pares o del procedimiento de análisis de varianza para experimentos con dos tratamientos en un diseño en bloque completamente aleatorio. Cuando es necesario investigar $k \geqslant 3$ tratamientos de un solo factor en presencia de un factor externo y por lo menos se encuentran disponibles mediciones ordinales, un método no paramétrico útil para determinar si los efectos debidos a los tratamientos son los mismos, es la *prueba de Friedman.*

De manera similar al procedimiento paramétrico, se crea un bloque para cada una de las n condiciones de los factores externos de tal manera que cada bloque contiene una observación proveniente de cada uno de los k tratamientos. Además, se supone que los tratamientos se asignan en forma aleatoria y que no existe ninguna interacción entre los bloques y los tratamientos. Las nk observaciones se arreglan como se ilustra en la tabla 15.4, donde los bloques son los renglones y los tratamientos las columnas.

La hipótesis nula para el procedimiento de Friedman es que los efectos atribuidos a los tratamientos son los mismos (es decir, las problaciones de interés tienen distribuciones idénticas) y la hipótesis alternativa es que existe una diferencia entre los tratamientos. Al igual que para la estadística de Kruskal-Wallis, las diferencias en los tratamientos descubiertas a través de la estadística de Friedman implican diferencias en la tendencia central.

TABLA 15.4 Arreglo de las observaciones para la prueba de Friedman

	Tratamiento					
Bloque	1	2	$\cdots$	j	$\cdots$	k
1	Y_{11}	Y_{12}	$\cdots$	Y_{1j}	$\cdots$	Y_{1k}
2	Y_{21}	Y_{22}	$\cdots$	Y_{2j}	$\cdots$	Y_{2k}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
n	Y_{n1}	Y_{n2}	$\cdots$	Y_{nj}	$\cdots$	Y_{nk}

Al igual que en los otros procedimientos no paramétricos, la prueba de Friedman se basa en los rangos. Para cada bloque (renglón) se asigna un rango a las observaciones comenzando con un rango 1 y terminando con un rango k; entonces se suman los rangos para cada tratamiento. Sea R_j la suma de los rangos del j-ésimo tratamiento (columna). Si dentro de cada bloque los efectos del tratamiento son los mismos, entonces para cualquier bloque los rangos deben ser una permutación aleatoria de los enteros del 1 al k, donde cada permutación tiene la misma probabilidad de ocurrencia. De esta forma, se espera que para cada tratamiento los rangos del 1 al k aparezcan, en forma aproximada, con la misma frecuencia. Si los efectos de los tratamientos son idénticos, R_j deberá tener prácticamente el mismo valor para toda j. Por lo tanto, el procedimiento de Friedman determina cuándo una disparidad observada entre los R_j es suficiente para rechazar la hipótesis nula.

La estadística de Friedman está dada por

$$S = \frac{12}{nk(k+1)}\left[\sum_{j=1}^{k} R_j^2\right] - 3n(k+1). \tag{15.14}$$

Las probabilidades para los valores de S se encuentran tabuladas para valores pequeños de n y k (véase [3]). Pero si el número de bloques n y el de tratamientos k no es muy pequeño (por ejemplo $n \geqslant 10$ y $k \geqslant 4$), la estadística S es, en forma aproximada, una variable aleatoria chi-cuadrada con $k - 1$ grados de libertad. Al igual que para la prueba de Kruskal-Wallis, la región crítica de tamaño α es la porción superior de la distribución chi-cuadrada con $k - 1$ grados de libertad. Se rechaza la hipótesis nula cuando el valor de S es mayor que el valor crítico. Al igual que en los casos anteriores, los empates se manejan mediante el uso de rangos promedio.

Ejemplo 15.5 Cuatro jueces se encargan de calificar en una competencia de salto que incluye a 10 finalistas. Los datos que figuran en la tabla 15.5 son las calificaciones, donde un 10 indica un salto perfecto. Para $\alpha = 0.01$, empléese la estadística de Friedman para determinar si existen diferencias discernibles en las calificaciones que otorgan cada uno de los cuatro jueces.

Los valores que figuran entre paréntesis en la tabla 15.5 son los rangos de las observaciones para cada competidor (bloque). Entonces, para cada juez, la suma de los

TABLA 15.5 Datos de la muestra para el ejemplo 15.5

	Juez			
Competidor	*1*	*2*	*3*	*4*
1	8.5 (3)	8.6 (4)	8.2 (1)	8.4 (2)
2	9.8 (4)	9.7 (3)	9.4 (1)	9.6 (2)
3	7.9 (2)	8.1 (3)	7.5 (1)	8.2 (4)
4	9.7 (3)	9.8 (4)	9.6 (1.5)	9.6 (1.5)
5	6.2 (1)	6.8 (3)	6.9 (4)	6.5 (2)
6	8.9 (3)	9.2 (4)	8.1 (1)	8.7 (2)
7	9.2 (3.5)	9.2 (3.5)	8.7 (1)	8.9 (2)
8	8.4 (1.5)	8.5 (3)	8.4 (1.5)	8.6 (4)
9	9.2 (2)	9.6 (4)	8.9 (1)	9.5 (3)
10	8.8 (2)	9.2 (3)	8.6 (1)	9.3 (4)

rangos es la siguiente: $R_1 = 25$, $R_2 = 34.5$, $R_3 = 14$, $R_4 = 26.5$. El valor de la estadística de Fiedman es

$$s = \frac{12}{(10)(4)(5)}[25^2 + 34.5^2 + 14^2 + 26.5^2] - (3)(10)(5) = 12.81.$$

Para $\alpha = 0.01$ y $k - 1 = 3$ grados de libertad, el valor crítico se obtiene de la tabla E del apéndice y es igual a 11.32. Dado que $s = 12.81 > 11.32$, se rechaza la hipótesis nula de que los efectos de los tratamientos son los mismos; las diferencias entre las calificaciones que otorgan los cuatro jueces son estadísticamente discernibles.

15.6 Coeficiente de correlación de rangos de Spearman

En la sección 13.8 se definió el coeficiente de correlación de la muestra como una medida de la asociación lineal que existe entre dos variables X y Y. El enfoque empleado en esa sección fue paramétrico, ya que se supuso una distribución normal bivariada para X y Y. En esta sección se define una popular medida no paramétrica de asociación cuando se emplean los rangos, que se conoce como coeficiente de correlación de rangos de Spearman, denotado por r_S.

Sean X y Y dos características de interés y supóngase que existe una muestra aleatoria de n pares que consiste sólo en los rangos de X y Y. El coeficiente de correlación del rango de Spearman es el coeficiente ordinario de correlación de la muestra que puede determinarse mediante el empleo, ya sea de (13.27) o (13.28), excepto que para este caso se emplean los rangos en lugar de las observaciones de X y Y. Al igual que el coeficiente de correlación de la muestra r, el coeficiente de correlación del rango r_S se define en el intervalo $-1 \leq r_S \leq 1$; y mide el grado de asociación lineal entre los rangos de X y Y. Para las características X y Y, la interpretación de r_S no es completamente idéntica a la de r. Si se tienen disponibles observaciones de X y Y, entonces el coeficiente de correlación de la muestra r es una medida del grado de asociación lineal que existe entre X y Y. Pero si se emplean los rangos, r_S mide la ten-

dencia de X y Y al relacionarse en forma monótona, es decir, r_S se encuentra cercano a 1 o a -1, se sugiere una asociación monótona creciente o decreciente para X y Y. En cierto sentido, r_S tiene un significado mayor que el de r debido a que al medir el grado de asociación monótona entre X y Y, r_S no se encuentra restringido a descubrir sólo una asociación lineal entre éstas.

Sea (X_i', Y_i'), $i = 1, 2, \ldots, n$ la representación de una muestra de rangos de X y Y. Entonces, de (13.28) el coeficiente de correlación de rangos de Spearman es

$$r_S = \frac{\sum_{i=1}^{n} X_i'Y_i' - \frac{\left(\sum_{i=1}^{n} X_i'\right)\left(\sum_{i=1}^{n} Y_i'\right)}{n}}{\left[\sum_{i=1}^{n} X_i'^2 - \frac{\left(\sum_{i=1}^{n} X_i'\right)^2}{n}\right]^{1/2}\left[\sum_{i=1}^{n} Y_i'^2 - \frac{\left(\sum_{i=1}^{n} Y_i'\right)^2}{n}\right]^{1/2}} \qquad (15.15)$$

Si no existen empates puede desarrollarse una relación alternativa más simple para (15.15) al tomar ventajas de la naturaleza del rango. Los rangos (X_i', Y_i') son arreglos de los primeros n enteros positivos. Dado que la suma de los primeros n enteros positivos es $n(n + 1)/2$, y la suma de sus cuadrados es $n(n + 1)(2n + 1)/6$,

$$\sum X_i' = \sum Y_i' = n(n + 1)/2 \qquad (15.16)$$

y

$$\sum X_i'^2 = \sum Y_i'^2 = n(n + 1)(2n + 1)/6. \qquad (15.17)$$

Además, dado que la relación

$$\sum X_i'Y_i' = \left[\sum X_i'^2 + \sum Y_i'^2 - \sum (X_i' - Y_i')^2\right]/2 \qquad (15.18)$$

es válida para cualquier valor, al sustituir (15.16) a (15.18) en (15.15) y después de algunos manejos algebráicos, se obtiene la expresión alternativa

$$r_S = 1 - \frac{6\sum(X_i' - Y_i')^2}{n(n^2 - 1)}. \qquad (15.19)$$

Ejemplo 15.6 Se pide a dos catadores de vinos que clasifiquen 10 vinos tintos ligeros en una escala del 1 (pobre) al 10 (excelente). Se obtienen los resultados que se muestran en la tabla 15.6. Calcúlese el coeficiente de correlación de rangos de Spearman.

Dado que no existen empates, puede usarse (15.19) para calcular r_S.

$$r_S = 1 - \frac{6[(5 - 3)^2 + (2 - 4)^2 + \cdots + (3 - 1)^2]}{10(100 - 1)} = 0.73,$$

lo cual sugiere una fuerte concordancia entre los dos catadores.

TABLA 15.6 Datos de la muestra para el ejemplo 15.6

Vino	*Catador 1* X'	*Catador 2* Y'
1	5	3
2	2	4
3	8	7
4	9	6
5	10	9
6	7	9
7	1	3
8	4	6
9	4	7
10	3	1

15.7 Comentarios finales

Para los métodos presentados en este capítulo, se tienen tres ventajas:

1. Las suposiciones para su empleo son menos estrictas que las de los correspondientes métodos paramétricos.
2. Los métodos no paramétricos pueden aplicarse en forma muy fácil a todas aquellas observaciones que se definen sobre una escala ordinal.
3. Los cálculos por efectuar son más fáciles cuando se comparan con los de los correspondientes métodos paramétricos.

A causa de la primera ventaja, los métodos paramétricos son particularmente útiles cuando se tienen muestras de tamaño pequeño y existe interés en adherirse a las suposiciones de distribución para los métodos paramétricos. En particular, las pruebas de Mann-Whitney, Wilcoxon, Kruskal-Wallis y de Friedman se comparan, en potencia, a las de los correspondientes métodos paramétricos, lo que incluye a la distribución t de Student o a la estadística F en el análisis de varianza, pero como ya se indicó en el capítulo 9, para muestras de tamaño mayor de 15, la distribución t de Student es bastante más robusta con respecto a la suposición de normalidad. Además, la estadística T es robusta con respecto a la suposición de varianzas iguales para muestras de gran tamaño y con el mismo número de observaciones, cuando se comparan dos medias, de la misma manera en que la estadística F lo es en el análisis de varianza, siempre y cuando los tamaños de la muestra de los tratamientos sean los mismos. De esta forma, cuando se tienen muestras de gran tamaño y las observaciones contenidas en éstas se definen por lo menos sobre una escala ordinal, puede perderse información muy importante al convertir las observaciones en rangos y signos y utilizar métodos no paramétricos. Para tales casos, la eficiencia en potencia de los métodos no paramétricos es menor que la de los procedimientos paramétricos. Por lo tanto, la ventaja más clara que tienen los métodos no paramétricos sobre los de tipo paramétrico es la segunda que se encuentra en la lista mencionada con anterioridad. La aplicación de los métodos paramétricos a observaciones que se en-

cuentran definidas sólo sobre una escala ordinal es muy difícil, ya que la interpretación de un intervalo en este caso tiene poco significado.

Referencias

1. J. D. Gibbons, *Nonparametric statistical inference*, McGraw-Hill, New York, 1971.
2. M. Hollander and D. A. Wolfe, *Nonparametric statistical methods*, Wiley, New York, 1973.
3. S. Siegel, *Nonparametric statistics for the behavioral sciences*, McGraw-Hill, New York, 1956.

Ejercicios

15.1. Para los datos del ejemplo 15.1, pruébese la hipótesis nula de que no existe diferencia entre las medias mediante el empleo del procedimiento t de Student de la sección 9.6.2. Para el mismo tamaño del error de tipo I dado en este ejemplo, ¿es diferente la conclusión?

15.2. Durante cinco años se llevó a cabo un estudio para determinar si existe alguna diferencia en el número de resfriados que sufren los fumadores y los no fumadores. Con base en muestras aleatorias de 14 no fumadores y 12 fumadores se observaron, a lo largo de los cinco años, los siguientes datos.

No fumadores	1	0	2	7	3	1	2	2	4	3	5	0	2	1
Fumadores	4	2	6	5	8	10	8	7	6	4	9	3		

Úsese la estadística U de Mann-Whitney para determinar si existe alguna razón para creer que estas muestras aleatorias provienen de poblaciones con diferentes distribuciones. Supóngase que $\alpha = 0.05$. ¿Existen algunas suposiciones que se estén violando?

15.3. Una compañía de mercadotecnia se interesa en comparar la aceptación por parte del consumidor de dos nuevos productos, A y B. Se seleccionaron, en forma aleatoria, 12 consumidores y se les pidió que dieran su opinión, con respecto al producto A, sobre una escala de 1 (Poca aceptación) a 5 (mucha aceptación). Se hizo lo mismo para el producto B, empleando para ello el mismo número de consumidores. Se obtuvieron los siguientes datos:

Producto A	1	2	5	5	4	3	5	4	4	3	5	2
Producto B	2	2	1	1	3	1	2	2	4	3	1	3

Mediante el empleo de la estadística U de Mann-Whitney, determínese si puede rechazarse, con $\alpha = 0.05$ la hipótesis nula de que estas muestras aleatorias provienen de poblaciones con distribuciones idénticas.

15.4. La siguiente información representa el número de unidades terminadas para dos trabajadores, A y B, en un periodo de cinco días.

A	49	52	53	47	50
B	56	48	58	46	55

a) Mediante el uso de la expresión (15.7), obténgase la función de probabilidad para el número de tendencias posible.

b) Para $\alpha = 0.05$, empléese el procedimiento de tendencias de Wald-Wolfowitz para probar la hipótesis nula de que estas muestras provienen de distribuciones idénticas.

15.5. El procedimiento de tendencias de Wald-Wolfowitz se emplea muchas veces para probar la aleatoriedad de una secuencia dada de observaciones. Si la aleatoriedad existe, entonces el número de tendencias para dos grupos distintos no deberá ser ni muy grande ni muy pequeño. Supóngase que los siguientes datos constituyen la secuencia de residuos para una ecuación de regresión estimada:

−2.98	−4.19	−0.51	5.19	2.38	6.73	0.93	1.29	−3.18
−1.14	−0.54	−2.76	−1.89	−4.28	−0.18	0.32	0.48	1.48
−2.43	−4.69	3.18	0.64	0.89	2.08	0.98	−3.28	

¿Existe alguna razón para creer que esta secuencia de residuos no es aleatoria? Úsese $\alpha = 0.05$.

15.6. Una compañía de mercadotecnia se interesa en la preferencia del consumidor con respecto a dos marcas de refresco que compiten entre sí. Se seleccionan, en forma aleatoria, 14 personas y se les pide que clasifiquen las bebidas mediante una escala del 1 (poca aceptación) al 10 (mucha aceptación). El orden en la selección de la bebida fue aleatorio. Se obtiene la siguiente información:

Persona	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*
Marca A	7	5	9	4	8	10	4	3	7	2	8	6	6	9
Marca B	3	2	7	6	9	3	5	1	4	2	4	7	5	4

Mediante el uso de la prueba del signo, ¿se tiene alguna razón para creer que existe una diferencia en la preferencia para estos dos refrescos? Supóngase $\alpha = 0.1$.

15.7. Para los datos que figuran en el ejercicio 15.6, empléese la prueba de rangos de signos de Wilcoxon. ¿Se obtienen las mismas conclusiones?

15.8. Para los datos del ejemplo 9.10, supóngase que no puede formularse la suposición de normalidad. Mediante el empleo de la prueba de rangos de signos de Wilcoxon, determínese si puede rechazarse la correspondiente hipótesis nula no paramétrica para un nivel $\alpha = 0.01$

15.9. Durante 12 días seleccionados al azar, dos tiendas, A y B vendieron el siguiente número de unidades del mismo producto:

Día	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
A	42	58	47	39	41	56	59	37	38	46	43	51
B	64	57	48	59	64	52	65	59	37	65	68	49

Mediante el empleo de la prueba del signo, ¿puede rechazarse la hipótesis nula de que las muestras provienen de distribuciones idénticas para un nivel $\alpha = 0.05$?

15.10. Para los datos que figuran en el ejercicio 15.9, úsese la prueba de rangos de signos de Wilcoxon y compárense los resultados.

15.11. Se desea determinar si el campo de especialización del estudiante no graduado tiene algún efecto sobre su desempeño en una escuela de leyes. Se toma una muestra aleatoria

TABLA 15.7 Datos de la muestra para el ejemplo 15.11

Finanzas	*Ciencia o ingeniería*	*Artes liberales*	*Otros*
9	3	2	14
22	7	4	34
24	10	15	48
31	18	26	52
47	23	38	59
65	25	43	63
		45	67
		49	72
		55	79

de 30 estudiantes de una clase de graduados de cierta escuela de leyes, la cual clasifica a los estudiantes y anota su campo de especialización; los datos que se encuentran en la tabla 15.7 son los resultados de este procedimiento. Mediante el empleo de la prueba de Kruskal-Wallis, determínese si el campo de especialización tiene algún efecto sobre el desempeño en la escuela de leyes, con $\alpha = 0.05$.

15.12. Con referencia a los datos que se encuentran en el ejercicio 12.7, empléese el procedimiento de Kruskal-Wallis para probar la hipótesis nula de que no existe ninguna diferencia con respecto a la durabilidad entre las dos marcas con $\alpha = 0.05$. La conclusión, ¿es la misma que la que se obtuvo para el ejercicio 12.7?

15.13. Se seleccionaron 12 estudiantes al azar, de una clase muy grande; sus calificaciones, en los cuatro exámenes que se llevaron a cabo durante el trimestre, se encuentran en la tabla 15.8. Mediante el uso de la prueba de Friedman, determínese si las diferencias entre los cuatro exámenes son estadísticamente discernibles para un nivel $\alpha = 0.01$. ¿Se estaría de acuerdo con la hipótesis de que no existe interacción alguna entre los estudiantes? Coméntese.

TABLA 15.8 Datos de la muestra para el ejercicio 15.13

Estudiante	*1*	*2*	*3*	*4*
1	72	68	80	75
2	89	87	78	92
3	48	56	64	58
4	65	76	70	62
5	86	94	93	85
6	56	73	78	87
7	75	84	65	69
8	39	45	48	56
9	78	67	69	59
10	98	87	86	95
11	64	87	92	48
12	82	76	85	79

15.14. Con referencia a los datos del ejercicio 12.6, úsese el procedimiento de Friedman para determinar si las diferencias que existen entre los cuatro supermercados son estadísticamente discernibles para un nivel $\alpha = 0.01$.

15.15. Para el ejercicio 13.12, conviértanse los datos en rangos y calcúlese el coeficiente de correlación de rangos de Spearman.

15.16. Dos jueces se encargan de calificar a ocho patinadores que patinan sobre hielo, mediante el empleo de una escala del 1 (muy malo) al 10 (el mejor). Se obtienen los siguiente resultados.

Patinador	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Juez 1	3	4	8	8	4	6	4	7
Juez 2	2	4	9	7	2	8	7	9

Calcúlese el coeficiente de correlación de rangos de Spearman y formúlese un comentario con respecto a si existe una relación que sea evidente.

15.17. Mediante el empleo de la misma escala que se menciona en el ejercicio 15.16, dos jueces califican el talento de las 10 semifinalistas del concurso señorita América. Se tienen los siguientes resultados:

Semifinalista	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Juez 1	2	6	5	9	3	7	9	2	6	2
Juez 2	7	1	4	4	8	9	3	9	10	8

Calcúlese el coeficiente de correlación de rangos de Spearman y formúlese un comentario con respecto a si existe alguna relación que sea evidente.

15.18. Un grupo de analistas de inversión clasificaron 10 compañías de acuerdo con su crecimiento potencial y el valor de sus acciones de la siguiente manera:

Compañía	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Valores en libros	8	3	10	1	6	2	5	7	4	9
Crecimiento	4	8	6	5	9	3	7	1	10	2

Calcúlese el coeficiente de correlación de rangos de Spearman y formúlese un comentario con respecto a si existe una relación, que sea evidente, entre el valor de las acciones de la compañía y su crecimiento potencial.

Apéndice

TABLA A	Valores de la función de distribución acumulativa binomial
TABLA B	Valores de la función de distribución acumulativa de Poisson
TABLA C	Valores de las funciones de probabilidad y distribución acumulativa para la distribución hipergeométrica
TABLA D	Valores de la función de distribución acumulativa normal estándar
TABLA E	Valores de cuantiles de la distribución chi-cuadrada
TABLA F	Valores de cuantiles de la distribución t de Student
TABLA G	Valores de cuantiles de la distribución F
TABLA H	k-valores para los límites de tolerancia bilaterales cuando se muestrean distribuciones normales
TABLA I	k-valores para los límites de tolerancia unilaterales cuando se muestrean distribuciones normales
TABLA J	Valores de cuantiles superiores de la distribución de la estadística D_n de Kolmogorov-Smirnov
TABLA K	Límites de la estadística de Durbin-Watson

TABLA A Valores de la función de distribución acumulativa binomial

$$P(X \le x) = F(x; n, p) = \sum_{i=0}^{x} \binom{n}{i} p^{x} (1-p)^{n-x}$$

		p										
n	x	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
2	0	0.9801	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
	1	0.9999	0.9975	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975	0.7500
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.9703	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
	1	0.9997	0.9928	0.9720	0.9392	0.8960	0.8438	0.7840	0.7183	0.6480	0.5748	0.5000
	2	1.0000	0.9999	0.9990	0.9966	0.9920	0.9844	0.9730	0.9571	0.9360	0.9089	0.8750
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.9606	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
	1	0.9994	0.9860	0.9477	0.8905	0.8192	0.7383	0.6517	0.5630	0.4752	0.3910	0.3125
	2	1.0000	0.9995	0.9963	0.9880	0.9728	0.9492	0.9163	0.8735	0.8208	0.7585	0.6875
	3	1.0000	1.0000	0.9999	0.9995	0.9984	0.9961	0.9919	0.9850	0.9744	0.9590	0.9375
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0	0.9510	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313
	1	0.9990	0.9774	0.9185	0.8352	0.7373	0.6328	0.5282	0.4284	0.3370	0.2562	0.1875
	2	1.0000	0.9988	0.9914	0.9734	0.9421	0.8965	0.8369	0.7648	0.6826	0.5931	0.5000
	3	1.0000	1.0000	0.9995	0.9978	0.9933	0.9844	0.9692	0.9460	0.9130	0.8688	0.8125
	4	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990	0.9976	0.9947	0.9898	0.9815	0.9688
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0	0.9415	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277	0.0156
	1	0.9985	0.9672	0.8857	0.7765	0.6554	0.5339	0.4202	0.3191	0.2333	0.1636	0.1094
	2	1.0000	0.9978	0.9842	0.9527	0.9011	0.8306	0.7443	0.6471	0.5443	0.4415	0.3438
	3	1.0000	0.9999	0.9987	0.9941	0.9830	0.9624	0.9295	0.8826	0.8208	0.7447	0.6563
	4	1.0000	1.0000	0.9999	0.9996	0.9984	0.9954	0.9891	0.9777	0.9590	0.9308	0.8906
	5	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9982	0.9959	0.9917	0.9844
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

7	0	0.9321	0.6983	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280	0.0152	0.0078
	1	0.9980	0.9556	0.8503	0.7166	0.5767	0.4449	0.3294	0.2338	0.1586	0.1024	0.0625
	2	1.0000	0.9962	0.9743	0.9262	0.8520	0.7564	0.6471	0.5323	0.4199	0.3164	0.2266
	3	1.0000	0.9998	0.9973	0.9879	0.9667	0.9294	0.8740	0.8002	0.7102	0.6083	0.5000
	4	1.0000	1.0000	0.9998	0.9988	0.9953	0.9871	0.9712	0.9444	0.9037	0.8471	0.7734
	5	1.0000	1.0000	1.0000	0.9999	0.9996	0.9987	0.9962	0.9910	0.9812	0.9643	0.9375
	6	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9984	0.9963	0.9922
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	0	0.9227	0.6634	0.4305	0.2725	0.1678	0.1001	0.0576	0.0319	0.0168	0.0084	0.0039
	1	0.9973	0.9428	0.8131	0.6572	0.5033	0.3671	0.2553	0.1691	0.1064	0.0632	0.0352
	2	0.9999	0.9942	0.9619	0.8948	0.7969	0.6785	0.5518	0.4278	0.3154	0.2201	0.1445
	3	1.0000	0.9996	0.9950	0.9786	0.9437	0.8862	0.8059	0.7064	0.5941	0.4770	0.3633
	4	1.0000	1.0000	0.9996	0.9971	0.9896	0.9727	0.9420	0.8939	0.8263	0.7396	0.6367
	5	1.0000	1.0000	1.0000	0.9998	0.9988	0.9958	0.9887	0.9747	0.9502	0.9115	0.8555
	6	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9987	0.9964	0.9915	0.9819	0.9648
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9983	0.9961
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	0	0.9135	0.6302	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101	0.0046	0.0020
	1	0.9966	0.9288	0.7748	0.5995	0.4362	0.3003	0.1960	0.1211	0.0705	0.0385	0.0195
	2	0.9999	0.9916	0.9470	0.8591	0.7382	0.6007	0.4628	0.3373	0.2318	0.1495	0.0898
	3	1.0000	0.9994	0.9917	0.9661	0.9144	0.8343	0.7297	0.6089	0.4826	0.3614	0.2539
	4	1.0000	1.0000	0.9991	0.9944	0.9804	0.9511	0.9012	0.8283	0.7334	0.6214	0.5000
	5	1.0000	1.0000	0.9999	0.9994	0.9969	0.9900	0.9747	0.9464	0.9006	0.8342	0.7461
	6	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	0.9957	0.9888	0.9750	0.9502	0.9102
	7	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9986	0.9962	0.9909	0.9805
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992	0.9980
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	0.9044	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010
	1	0.9957	0.9139	0.7361	0.5443	0.3758	0.2440	0.1495	0.0860	0.0464	0.0233	0.0107
	2	0.9999	0.9885	0.9298	0.8202	0.6778	0.5256	0.3828	0.2616	0.1673	0.0996	0.0547
	3	1.0000	0.9990	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3823	0.2660	0.1719
	4	1.0000	0.9999	0.9984	0.9901	0.9672	0.9219	0.8497	0.7515	0.6331	0.5044	0.3770
	5	1.0000	1.0000	0.9999	0.9986	0.9936	0.9803	0.9527	0.9051	0.8338	0.7384	0.6230
	6	1.0000	1.0000	1.0000	0.9999	0.9991	0.9965	0.9894	0.9740	0.9452	0.8980	0.8281
	7	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9952	0.9877	0.9726	0.9453
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983	0.9955	0.9893
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

TABLA A (continuación) Valores de la función de distribución acumulativa binomial

		p										
n	x	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
11	0	0.8953	0.5688	0.3138	0.1673	0.0859	0.0422	0.0198	0.0088	0.0036	0.0014	0.0005
	1	0.9948	0.8981	0.6974	0.4922	0.3221	0.1971	0.1130	0.0606	0.0302	0.0139	0.0059
	2	0.9998	0.9848	0.9104	0.7788	0.6174	0.4552	0.3127	0.2001	0.1189	0.0652	0.0327
	3	1.0000	0.9984	0.9815	0.9306	0.8389	0.7133	0.5696	0.4256	0.2963	0.1911	0.1133
	4	1.0000	0.9999	0.9972	0.9841	0.9496	0.8854	0.7897	0.6683	0.5328	0.3971	0.2744
	5	1.0000	1.0000	0.9997	0.9973	0.9883	0.9657	0.9218	0.8513	0.7535	0.6331	0.5000
	6	1.0000	1.0000	1.0000	0.9997	0.9980	0.9924	0.9784	0.9499	0.9006	0.8262	0.7256
	7	1.0000	1.0000	1.0000	1.0000	0.9998	0.9988	0.9957	0.9878	0.9707	0.9390	0.8867
	8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9980	0.9941	0.9852	0.9673
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9993	0.9978	0.9941
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9995
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
12	0	0.8864	0.5404	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057	0.0022	0.0008	0.0002
	1	0.9938	0.8816	0.6590	0.4435	0.2749	0.1584	0.0850	0.0424	0.0196	0.0083	0.0032
	2	0.9998	0.9804	0.8891	0.7358	0.5583	0.3907	0.2528	0.1513	0.0834	0.0421	0.0193
	3	1.0000	0.9978	0.9744	0.9078	0.7946	0.6488	0.4925	0.3467	0.2253	0.1345	0.0730
	4	1.0000	0.9998	0.9957	0.9761	0.9274	0.8424	0.7237	0.5833	0.4382	0.3044	0.1938
	5	1.0000	1.0000	0.9995	0.9954	0.9806	0.9456	0.8822	0.7873	0.6652	0.5269	0.3872
	6	1.0000	1.0000	0.9999	0.9993	0.9961	0.9857	0.9614	0.9154	0.8418	0.7393	0.6128
	7	1.0000	1.0000	1.0000	0.9999	0.9994	0.9972	0.9905	0.9745	0.9427	0.8883	0.8062
	8	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9983	0.9944	0.9847	0.9644	0.9270
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9992	0.9972	0.9921	0.9807
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9968
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

13	0	0.8775	0.5133	0.2542	0.1209	0.0550	0.0238	0.0097	0.0037	0.0013	0.0004	0.0001
	1	0.9928	0.8646	0.6213	0.3983	0.2336	0.1267	0.0637	0.0296	0.0126	0.0049	0.0017
	2	0.9997	0.9755	0.8661	0.6920	0.5017	0.3326	0.2025	0.1132	0.0579	0.0269	0.0112
	3	1.0000	0.9969	0.9658	0.8820	0.7473	0.5843	0.4206	0.2783	0.1686	0.0929	0.0461
	4	1.0000	0.9997	0.9935	0.9658	0.9009	0.7940	0.6543	0.5005	0.3530	0.2279	0.1334
	5	1.0000	1.0000	0.9991	0.9925	0.9700	0.9198	0.8346	0.7159	0.5744	0.4268	0.2905
	6	1.0000	1.0000	0.9999	0.9987	0.9930	0.9757	0.9376	0.8705	0.7712	0.6437	0.5000
	7	1.0000	1.0000	1.0000	0.9998	0.9988	0.9944	0.9818	0.9538	0.9023	0.8212	0.7095
	8	1.0000	1.0000	1.0000	1.0000	0.9998	0.9990	0.9960	0.9874	0.9679	0.9302	0.8666
	9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9975	0.9922	0.9797	0.9539
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9987	0.9959	0.9888
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
14	0	0.8687	0.4877	0.2288	0.1028	0.0440	0.0178	0.0068	0.0024	0.0008	0.0002	0.0001
	1	0.9916	0.8470	0.5846	0.3567	0.1979	0.1010	0.0475	0.0205	0.0081	0.0029	0.0009
	2	0.9997	0.9699	0.8416	0.6479	0.4481	0.2811	0.1608	0.0839	0.0398	0.0170	0.0065
	3	1.0000	0.9958	0.9559	0.8535	0.6982	0.5213	0.3552	0.2205	0.1243	0.0632	0.0287
	4	1.0000	0.9996	0.9908	0.9533	0.8702	0.7415	0.5842	0.4227	0.2793	0.1672	0.0898
	5	1.0000	1.0000	0.9985	0.9885	0.9561	0.8883	0.7805	0.6405	0.4859	0.3373	0.2120
	6	1.0000	1.0000	0.9998	0.9978	0.9884	0.9617	0.9067	0.8164	0.6925	0.5461	0.3953
	7	1.0000	1.0000	1.0000	0.9997	0.9976	0.9897	0.9685	0.9247	0.8499	0.7414	0.6047
	8	1.0000	1.0000	1.0000	1.0000	0.9996	0.9978	0.9917	0.9757	0.9417	0.8811	0.7880
	9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9983	0.9940	0.9825	0.9574	0.9102
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9989	0.9961	0.9886	0.9713
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9978	0.9935
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9991
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

TABLA A (continuación) Valores de la función de distribución acumulativa binomial

		p										
n	x	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
15	0	0.8601	0.4633	0.2059	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001	0.0000
	1	0.9904	0.8290	0.5490	0.3186	0.1671	0.0802	0.0353	0.0142	0.0052	0.0017	0.0005
	2	0.9996	0.9638	0.8159	0.6042	0.3980	0.2361	0.1268	0.0617	0.0271	0.0107	0.0037
	3	1.0000	0.9945	0.9444	0.8227	0.6482	0.4613	0.2969	0.1727	0.0905	0.0424	0.0176
	4	1.0000	0.9994	0.9873	0.9383	0.8358	0.6865	0.5155	0.3519	0.2173	0.1204	0.0592
	5	1.0000	0.9999	0.9978	0.9832	0.9389	0.8516	0.7216	0.5643	0.4032	0.2608	0.1509
	6	1.0000	1.0000	0.9997	0.9964	0.9819	0.9434	0.8689	0.7548	0.6098	0.4522	0.3036
	7	1.0000	1.0000	1.0000	0.9994	0.9958	0.9827	0.9500	0.8868	0.7869	0.6535	0.5000
	8	1.0000	1.0000	1.0000	0.9999	0.9992	0.9958	0.9848	0.9578	0.9050	0.8182	0.6964
	9	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9963	0.9876	0.9662	0.9231	0.8491
	10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9972	0.9907	0.9745	0.9408
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9937	0.9824
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9963
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
16	0	0.8515	0.4401	0.1853	0.0743	0.0281	0.0100	0.0033	0.0010	0.0003	0.0001	0.0000
	1	0.9891	0.8108	0.5147	0.2839	0.1407	0.0635	0.0261	0.0098	0.0033	0.0010	0.0003
	2	0.9995	0.9571	0.7893	0.5614	0.3518	0.1971	0.0994	0.0451	0.0183	0.0066	0.0021
	3	1.0000	0.9930	0.9316	0.7899	0.5981	0.4050	0.2459	0.1339	0.0651	0.0281	0.0106
	4	1.0000	0.9991	0.9830	0.9209	0.7982	0.6302	0.4499	0.2892	0.1666	0.0853	0.0384
	5	1.0000	0.9999	0.9967	0.9765	0.9183	0.8103	0.6598	0.4900	0.3288	0.1976	0.1051
	6	1.0000	1.0000	0.9995	0.9944	0.9733	0.9204	0.8247	0.6881	0.5272	0.3660	0.2272
	7	1.0000	1.0000	0.9999	0.9989	0.9930	0.9729	0.9256	0.8406	0.7161	0.5629	0.4018
	8	1.0000	1.0000	1.0000	0.9998	0.9985	0.9925	0.9743	0.9329	0.8577	0.7441	0.5982
	9	1.0000	1.0000	1.0000	1.0000	0.9998	0.9984	0.9929	0.9771	0.9417	0.8759	0.7728
	10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9938	0.9809	0.9514	0.8949
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	0.9951	0.9851	0.9616
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9991	0.9965	0.9894
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9979
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

17	0	0.8429	0.4181	0.1668	0.0631	0.0225	0.0075	0.0023	0.0007	0.0002	0.0000	0.0000
	1	0.9877	0.7922	0.4818	0.2525	0.1182	0.0501	0.0193	0.0067	0.0021	0.0006	0.0001
	2	0.9994	0.9497	0.7618	0.5198	0.3096	0.1637	0.0774	0.0327	0.0123	0.0041	0.0012
	3	1.0000	0.9912	0.9174	0.7556	0.5489	0.3530	0.2019	0.1028	0.0464	0.0184	0.0064
	4	1.0000	0.9988	0.9779	0.9013	0.7582	0.5739	0.3887	0.2348	0.1260	0.0596	0.0245
	5	1.0000	0.9999	0.9953	0.9681	0.8943	0.7653	0.5968	0.4197	0.2639	0.1471	0.0717
	6	1.0000	1.0000	0.9992	0.9917	0.9623	0.8929	0.7752	0.6188	0.4478	0.2902	0.1662
	7	1.0000	1.0000	0.9999	0.9983	0.9891	0.9598	0.8954	0.7872	0.6405	0.4743	0.3145
	8	1.0000	1.0000	1.0000	0.9997	0.9974	0.9876	0.9597	0.9006	0.8011	0.6626	0.5000
	9	1.0000	1.0000	1.0000	1.0000	0.9995	0.9969	0.9873	0.9617	0.9081	0.8166	0.6855
	10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9968	0.9880	0.9652	0.9174	0.8338
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9970	0.9894	0.9699	0.9283
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9975	0.9914	0.9755
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9936
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9988
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
18	0	0.8345	0.3972	0.1501	0.0536	0.0180	0.0056	0.0016	0.0004	0.0001	0.0000	0.0000
	1	0.9862	0.7735	0.4503	0.2241	0.0991	0.0395	0.0142	0.0046	0.0013	0.0003	0.0001
	2	0.9993	0.9419	0.7338	0.4797	0.2713	0.1353	0.0600	0.0236	0.0082	0.0025	0.0007
	3	1.0000	0.9891	0.9018	0.7202	0.5010	0.3057	0.1646	0.0783	0.0328	0.0120	0.0038
	4	1.0000	0.9985	0.9718	0.8794	0.7164	0.5187	0.3327	0.1886	0.0942	0.0411	0.0154
	5	1.0000	0.9998	0.9936	0.9581	0.8671	0.7175	0.5344	0.3550	0.2088	0.1077	0.0481
	6	1.0000	1.0000	0.9988	0.9882	0.9487	0.8610	0.7217	0.5491	0.3743	0.2258	0.1189
	7	1.0000	1.0000	0.9998	0.9973	0.9837	0.9431	0.8593	0.7283	0.5634	0.3915	0.2403
	8	1.0000	1.0000	1.0000	0.9995	0.9957	0.9807	0.9404	0.8609	0.7368	0.5778	0.4073
	9	1.0000	1.0000	1.0000	0.9999	0.9991	0.9946	0.9790	0.9403	0.8653	0.7473	0.5927
	10	1.0000	1.0000	1.0000	1.0000	0.9998	0.9988	0.9939	0.9788	0.9424	0.8720	0.7597
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9986	0.9938	0.9797	0.9463	0.8811
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9986	0.9942	0.9817	0.9519
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	0.9951	0.9846
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9990	0.9962
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

TABLA A (continuación) Valores de la función de distribución acumulativa binomial

		p										
n	x	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
19	0	0.8262	0.3774	0.1351	0.0456	0.0144	0.0042	0.0011	0.0003	0.0001	0.0000	0.0000
	1	0.9847	0.7547	0.4203	0.1985	0.0829	0.0310	0.0104	0.0031	0.0008	0.0002	0.0000
	2	0.9991	0.9335	0.7054	0.4413	0.2369	0.1113	0.0462	0.0170	0.0055	0.0015	0.0004
	3	1.0000	0.9868	0.8850	0.6841	0.4551	0.2631	0.1332	0.0591	0.0230	0.0077	0.0022
	4	1.0000	0.9980	0.9648	0.8556	0.6733	0.4654	0.2822	0.1500	0.0696	0.0280	0.0096
	5	1.0000	0.9998	0.9914	0.9463	0.8369	0.6678	0.4739	0.2968	0.1629	0.0777	0.0318
	6	1.0000	1.0000	0.9983	0.9837	0.9324	0.8251	0.6655	0.4812	0.3081	0.1727	0.0835
	7	1.0000	1.0000	0.9997	0.9959	0.9767	0.9225	0.8180	0.6656	0.4878	0.3169	0.1796
	8	1.0000	1.0000	1.0000	0.9992	0.9933	0.9713	0.9161	0.8145	0.6675	0.4940	0.3238
	9	1.0000	1.0000	1.0000	0.9999	0.9984	0.9911	0.9674	0.9125	0.8139	0.6710	0.5000
	10	1.0000	1.0000	1.0000	1.0000	0.9997	0.9977	0.9895	0.9653	0.9115	0.8159	0.6762
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9972	0.9886	0.9648	0.9129	0.8204
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9969	0.9884	0.9658	0.9165
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9969	0.9891	0.9682
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9972	0.9904
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9978
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0	0.8179	0.3585	0.1216	0.0388	0.0115	0.0032	0.0008	0.0002	0.0000	0.0000	0.0000
	1	0.9831	0.7358	0.3917	0.1756	0.0692	0.0243	0.0076	0.0021	0.0005	0.0001	0.0000
	2	0.9990	0.9245	0.6769	0.4049	0.2061	0.0913	0.0355	0.0121	0.0036	0.0009	0.0002
	3	1.0000	0.9841	0.8670	0.6477	0.4114	0.2252	0.1071	0.0444	0.0160	0.0049	0.0013
	4	1.0000	0.9974	0.9568	0.8298	0.6296	0.4148	0.2375	0.1182	0.0510	0.0189	0.0059
	5	1.0000	0.9997	0.9887	0.9327	0.8042	0.6172	0.4164	0.2454	0.1256	0.0553	0.0207
	6	1.0000	1.0000	0.9976	0.9781	0.9133	0.7858	0.6080	0.4166	0.2500	0.1299	0.0577
	7	1.0000	1.0000	0.9996	0.9941	0.9679	0.8982	0.7723	0.6010	0.4159	0.2520	0.1316
	8	1.0000	1.0000	0.9999	0.9987	0.9900	0.9591	0.8867	0.7624	0.5956	0.4143	0.2517
	9	1.0000	1.0000	1.0000	0.9998	0.9974	0.9861	0.9520	0.8782	0.7553	0.5914	0.4119
	10	1.0000	1.0000	1.0000	1.0000	0.9994	0.9961	0.9829	0.9468	0.8725	0.7507	0.5881

	11	1.0000	1.0000	1.0000	1.0000	0.9999	0.9991	0.9949	0.9804	0.9435	0.8692	0,7483
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9987	0.9940	0.9790	0.9420	0.8684
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9935	0.9786	0.9423
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9936	0.9793
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9941
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
25	0	0.7778	0.2774	0.0718	0.0172	0.0038	0.0008	0.0001	0.0000	0.0000	0.0000	0.0000
	1	0.9742	0.6424	0.2712	0.0931	0.0274	0.0070	0.0016	0.0003	0.0001	0.0000	0.0000
	2	0.9980	0.8729	0.5371	0.2537	0.0982	0.0321	0.0090	0.0021	0.0004	0.0001	0.0000
	3	0.9999	0.9659	0.7636	0.4711	0.2340	0.0962	0.0332	0.0097	0.0024	0.0005	0.0001
	4	1.0000	0.9928	0.9020	0.6821	0.4207	0.2137	0.0905	0.0320	0.0095	0.0023	0.0005
	5	1.0000	0.9988	0.9666	0.8385	0.6167	0.3783	0.1935	0.0826	0.0294	0.0086	0.0020
	6	1.0000	0.9998	0.9905	0.9305	0.7800	0.5611	0.3407	0.1734	0.0736	0.0258	0.0073
	7	1.0000	1.0000	0.9977	0.9745	0.8909	0.7265	0.5118	0.3061	0.1536	0.0639	0.0216
	8	1.0000	1.0000	0.9995	0.9920	0.9532	0.8506	0.6769	0.4668	0.2735	0.1340	0.0539
	9	1.0000	1.0000	0.9999	0.9979	0.9827	0.9287	0.8106	0.6303	0.4246	0.2424	0.1148
	10	1.0000	1.0000	1.0000	0.9995	0.9944	0.9703	0.9022	0.7712	0.5858	0.3843	0.2122
	11	1.0000	1.0000	1.0000	0.9999	0.9985	0.9893	0.9558	0.8746	0.7323	0.5426	0.3450
	12	1.0000	1.0000	1.0000	1.0000	0.9996	0.9966	0.9825	0.9396	0.8462	0.6937	0.5000
	13	1.0000	1.0000	1.0000	1.0000	0.9999	0.9991	0.9940	0.9745	0.9222	0.8173	0.6550
	14	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9982	0.9907	0.9656	0.9040	0.7878
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9971	0.9868	0.9560	0.8852
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9957	0.9826	0.9461
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9988	0.9942	0.9784
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9927
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9980
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995
	21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

TABLA B Valores de la función de distribución acumulativa de Poisson

$$P(X \leq x) = F(x; \lambda) = \sum_{i=0}^{x} \frac{e^{-\lambda}\lambda^i}{i!}$$

	λ									
x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	0.3679
1	0.9953	0.9825	0.9631	0.9384	0.9098	0.8781	0.8442	0.8088	0.7725	0.7358
2	0.9998	0.9989	0.9964	0.9921	0.9856	0.9769	0.9659	0.9526	0.9371	0.9197
3	1.0000	0.9999	0.9997	0.9992	0.9982	0.9966	0.99[illegible]2	0.9909	0.9865	0.9810
4	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9992	0.9986	0.9977	0.9963
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9994
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

	λ									
x	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0	0.3329	0.3012	0.2725	0.2466	0.2231	0.2019	0.1827	0.1653	0.1496	0.1353
1	0.6990	0.6626	0.6268	0.5918	0.5578	0.5249	0.4932	0.4628	0.4338	0.4060
2	0.9004	0.8795	0.8571	0.8335	0.8088	0.7834	0.7572	0.7306	0.7037	0.6767
3	0.9743	0.9662	0.9569	0.9463	0.9344	0.9212	0.9068	0.8913	0.8747	0.8571
4	0.9946	0.9923	0.9893	0.9857	0.9814	0.9763	0.9704	0.9636	0.9559	0.9473
5	0.9990	0.9985	0.9978	0.9968	0.9955	0.9940	0.9920	0.9896	0.9868	0.9834
6	0.9999	0.9997	0.9996	0.9994	0.9991	0.9987	0.9981	0.9974	0.9966	0.9955
7	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9996	0.9994	0.9992	0.9989
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9998
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	λ = 2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
0	0.1225	0.1108	0.1003	0.0907	0.0821	0.0743	0.0672	0.0608	0.0550	0.0498
1	0.3796	0.3546	0.3309	0.3084	0.2873	0.2674	0.2487	0.2311	0.2146	0.1991
2	0.6496	0.6227	0.5960	0.5697	0.5438	0.5184	0.4936	0.4695	0.4460	0.4232
3	0.8386	0.8194	0.7993	0.7787	0.7576	0.7360	0.7141	0.6919	0.6696	0.6472
4	0.9379	0.9275	0.9163	0.9041	0.8912	0.8774	0.8629	0.8477	0.8318	0.8153
5	0.9796	0.9751	0.9700	0.9643	0.9580	0.9510	0.9433	0.9349	0.9258	0.9161
6	0.9941	0.9925	0.9906	0.9884	0.9858	0.9828	0.9794	0.9756	0.9713	0.9665
7	0.9985	0.9980	0.9974	0.9967	0.9958	0.9947	0.9934	0.9919	0.9901	0.9881
8	0.9997	0.9995	0.9994	0.9991	0.9989	0.9985	0.9981	0.9976	0.9969	0.9962
9	0.9999	0.9999	0.9999	0.9998	0.9997	0.9996	0.9995	0.9993	0.9991	0.9989
10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9998	0.9998	0.9997
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999

x	λ = 3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
0	0.0450	0.0408	0.0369	0.0334	0.0302	0.0273	0.0247	0.0224	0.0202	0.0183
1	0.1847	0.1712	0.1586	0.1468	0.1359	0.1257	0.1162	0.1074	0.0992	0.0916
2	0.4012	0.3799	0.3594	0.3397	0.3208	0.3027	0.2854	0.2689	0.2531	0.2381
3	0.6248	0.6025	0.5803	0.5584	0.5366	0.5152	0.4942	0.4735	0.4533	0.4335
4	0.7982	0.7806	0.7626	0.7442	0.7254	0.7064	0.6872	0.6678	0.6484	0.6288
5	0.9057	0.8946	0.8829	0.8705	0.8576	0.8441	0.8301	0.8156	0.8006	0.7851
6	0.9612	0.9554	0.9490	0.9421	0.9347	0.9267	0.9182	0.9091	0.8995	0.8893
7	0.9858	0.9832	0.9802	0.9769	0.9733	0.9692	0.9648	0.9599	0.9546	0.9489
8	0.9953	0.9943	0.9931	0.9917	0.9901	0.9883	0.9863	0.9840	0.9815	0.9786
9	0.9986	0.9982	0.9978	0.9973	0.9967	0.9960	0.9952	0.9942	0.9931	0.9919
10	0.9996	0.9995	0.9994	0.9992	0.9990	0.9987	0.9984	0.9981	0.9977	0.9972
11	0.9999	0.9999	0.9998	0.9998	0.9997	0.9996	0.9995	0.9994	0.9993	0.9991
12	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998	0.9997
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999

TABLA B (continuación) Valores de la función de distribución acumulativa de Poisson

	λ									
x	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0
0	0.0166	0.0150	0.0136	0.0123	0.0111	0.0101	0.0091	0.0082	0.0074	0.0067
1	0.0845	0.0780	0.0719	0.0663	0.0611	0.0563	0.0518	0.0477	0.0439	0.0404
2	0.2238	0.2102	0.1974	0.1851	0.1736	0.1626	0.1523	0.1425	0.1333	0.1247
3	0.4142	0.3954	0.3772	0.3595	0.3423	0.3257	0.3097	0.2942	0.2793	0.2650
4	0.6093	0.5898	0.5704	0.5512	0.5321	0.5132	0.4946	0.4763	0.4582	0.4405
5	0.7693	0.7531	0.7367	0.7199	0.7029	0.6858	0.6684	0.6510	0.6335	0.6160
6	0.8787	0.8675	0.8558	0.8436	0.8311	0.8180	0.8046	0.7908	0.7767	0.7622
7	0.9427	0.9361	0.9290	0.9214	0.9134	0.9050	0.8960	0.8867	0.8769	0.8666
8	0.9755	0.9721	0.9683	0.9642	0.9597	0.9549	0.9497	0.9442	0.9382	0.9319
9	0.9905	0.9889	0.9871	0.9851	0.9829	0.9805	0.9778	0.9749	0.9717	0.9682
10	0.9966	0.9959	0.9952	0.9943	0.9933	0.9922	0.9910	0.9896	0.9880	0.9863
11	0.9989	0.9986	0.9983	0.9980	0.9976	0.9971	0.9966	0.9960	0.9953	0.9945
12	0.9997	0.9996	0.9995	0.9993	0.9992	0.9990	0.9988	0.9986	0.9983	0.9980
13	0.9999	0.9999	0.9998	0.9998	0.9997	0.9997	0.9996	0.9995	0.9994	0.9993
14	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999

	λ									
x	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.0
0	0.0061	0.0055	0.0050	0.0045	0.0041	0.0037	0.0033	0.0030	0.0027	0.0025
1	0.0372	0.0342	0.0314	0.0289	0.0266	0.0244	0.0224	0.0206	0.0189	0.0174
2	0.1165	0.1088	0.1016	0.0948	0.0884	0.0824	0.0768	0.0715	0.0666	0.0620
3	0.2513	0.2381	0.2254	0.2133	0.2017	0.1906	0.1801	0.1700	0.1604	0.1512
4	0.4231	0.4061	0.3895	0.3733	0.3575	0.3422	0.3272	0.3127	0.2987	0.2851
5	0.5984	0.5809	0.5635	0.5461	0.5289	0.5119	0.4950	0.4783	0.4619	0.4457
6	0.7474	0.7324	0.7171	0.7017	0.6860	0.6703	0.6544	0.6384	0.6224	0.6063
7	0.8560	0.8449	0.8335	0.8217	0.8095	0.7970	0.7842	0.7710	0.7576	0.7440
8	0.9252	0.9181	0.9106	0.9027	0.8944	0.8857	0.8766	0.8672	0.8574	0.8472
9	0.9644	0.9603	0.9559	0.9512	0.9462	0.9409	0.9352	0.9292	0.9228	0.9161
10	0.9844	0.9823	0.9800	0.9775	0.9747	0.9718	0.9686	0.9651	0.9614	0.9574
11	0.9937	0.9927	0.9916	0.9904	0.9890	0.9875	0.9859	0.9841	0.9821	0.9799
12	0.9976	0.9972	0.9967	0.9962	0.9955	0.9949	0.9941	0.9932	0.9922	0.9912
13	0.9992	0.9990	0.9988	0.9986	0.9983	0.9980	0.9977	0.9973	0.9969	0.9964
14	0.9997	0.9997	0.9996	0.9995	0.9994	0.9993	0.9991	0.9990	0.9988	0.9986
15	0.9999	0.9999	0.9999	0.9998	0.9998	0.9998	0.9997	0.9996	0.9996	0.9995
16	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999

	λ									
x	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	7.0
0	0.0022	0.0020	0.0018	0.0017	0.0015	0.0014	0.0012	0.0011	0.0010	0.0009
1	0.0159	0.0146	0.0134	0.0123	0.0113	0.0103	0.0095	0.0087	0.0080	0.0073
2	0.0577	0.0536	0.0498	0.0463	0.0430	0.0400	0.0371	0.0344	0.0320	0.0296
3	0.1425	0.1342	0.1264	0.1189	0.1119	0.1052	0.0988	0.0928	0.0871	0.0818
4	0.2719	0.2592	0.2469	0.2351	0.2237	0.2127	0.2022	0.1920	0.1823	0.1730
5	0.4298	0.4141	0.3988	0.3837	0.3690	0.3547	0.3407	0.3270	0.3137	0.3007
6	0.5902	0.5742	0.5582	0.5423	0.5265	0.5108	0.4953	0.4799	0.4647	0.4497
7	0.7301	0.7160	0.7018	0.6873	0.6728	0.6581	0.6433	0.6285	0.6136	0.5987
8	0.8367	0.8259	0.8148	0.8033	0.7916	0.7796	0.7673	0.7548	0.7420	0.7291
9	0.9090	0.9016	0.8939	0.8858	0.8774	0.8686	0.8596	0.8502	0.8405	0.8305
10	0.9531	0.9486	0.9437	0.9386	0.9332	0.9274	0.9214	0.9151	0.9084	0.9015
11	0.9776	0.9750	0.9723	0.9693	0.9661	0.9627	0.9591	0.9552	0.9510	0.9467
12	0.9900	0.9887	0.9873	0.9857	0.9840	0.9821	0.9801	0.9779	0.9755	0.9730
13	0.9958	0.9952	0.9945	0.9937	0.9929	0.9920	0.9909	0.9898	0.9885	0.9872
14	0.9984	0.9981	0.9978	0.9974	0.9970	0.9966	0.9961	0.9956	0.9950	0.9943
15	0.9994	0.9993	0.9992	0.9990	0.9988	0.9986	0.9984	0.9982	0.9979	0.9976
16	0.9998	0.9997	0.9997	0.9996	0.9996	0.9995	0.9994	0.9993	0.9992	0.9990
17	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998	0.9998	0.9997	0.9997	0.9996
18	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

TABLA B (continuación) Valores de la función de distribución acumulativa de Poisson

	λ									
x	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	8.0
0	0.0008	0.0007	0.0007	0.0006	0.0006	0.0005	0.0005	0.0004	0.0004	0.0003
1	0.0067	0.0061	0.0056	0.0051	0.0047	0.0043	0.0039	0.0036	0.0033	0.0030
2	0.0275	0.0255	0.0236	0.0219	0.0203	0.0188	0.0174	0.0161	0.0149	0.0138
3	0.0767	0.0719	0.0674	0.0632	0.0591	0.0554	0.0518	0.0485	0.0453	0.0424
4	0.1641	0.1555	0.1473	0.1395	0.1321	0.1249	0.1181	0.1117	0.1055	0.0996
5	0.2881	0.2759	0.2640	0.2526	0.2414	0.2307	0.2203	0.2103	0.2006	0.1912
6	0.4349	0.4204	0.4060	0.3920	0.3782	0.3646	0.3514	0.3384	0.3257	0.3134
7	0.5838	0.5689	0.5541	0.5393	0.5246	0.5100	0.4956	0.4812	0.4670	0.4530
8	0.7160	0.7027	0.6892	0.6757	0.6620	0.6482	0.6343	0.6204	0.6065	0.5926
9	0.8202	0.8097	0.7988	0.7877	0.7764	0.7649	0.7531	0.7411	0.7290	0.7166
10	0.8942	0.8867	0.8788	0.8707	0.8622	0.8535	0.8445	0.8352	0.8257	0.8159
11	0.9420	0.9371	0.9319	0.9265	0.9208	0.9148	0.9085	0.9020	0.8952	0.8881
12	0.9703	0.9673	0.9642	0.9609	0.9573	0.9536	0.9496	0.9454	0.9409	0.9362
13	0.9857	0.9841	0.9824	0.9805	0.9784	0.9762	0.9739	0.9714	0.9687	0.9658
14	0.9935	0.9927	0.9918	0.9908	0.9897	0.9886	0.9873	0.9859	0.9844	0.9827
15	0.9972	0.9969	0.9964	0.9959	0.9954	0.9948	0.9941	0.9934	0.9926	0.9918
16	0.9989	0.9987	0.9985	0.9983	0.9980	0.9978	0.9974	0.9971	0.9967	0.9963
17	0.9996	0.9995	0.9994	0.9993	0.9992	0.9991	0.9989	0.9988	0.9986	0.9984
18	0.9998	0.9998	0.9998	0.9997	0.9997	0.9996	0.9996	0.9995	0.9994	0.9993
19	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998	0.9998	0.9997
20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999
21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	λ = 8.1	8.2	8.3	8.4	8.5	8.6	8.7	8.8	8.9	9.0
0	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001
1	0.0028	0.0025	0.0023	0.0021	0.0019	0.0018	0.0016	0.0015	0.0014	0.0012
2	0.0127	0.0118	0.0109	0.0100	0.0093	0.0086	0.0079	0.0073	0.0068	0.0062
3	0.0396	0.0370	0.0346	0.0323	0.0301	0.0281	0.0262	0.0244	0.0228	0.0212
4	0.0941	0.0887	0.0837	0.0789	0.0744	0.0701	0.0660	0.0621	0.0584	0.0550
5	0.1823	0.1736	0.1653	0.1573	0.1496	0.1422	0 1352	0.1284	0.1219	0.1157
6	0.3013	0.2896	0.2781	0.2670	0.2562	0.2457	0.2355	0.2256	0.2160	0.2068
7	0.4391	0.4254	0.4119	0.3987	0.3856	0.3728	0.3602	0.3478	0.3357	0.3239
8	0.5786	0.5647	0.5508	0.5369	0.5231	0.5094	0.4958	0.4823	0.4689	0.4557
9	0.7041	0.6915	0.6788	0.6659	0.6530	0.6400	0.6269	0.6137	0.6006	0.5874
10	0.8058	0.7956	0.7850	0.7743	0.7634	0.7522	0.7409	0.7294	0.7178	0.7060
11	0.8807	0.8731	0.8652	0.8571	0.8487	0.8400	0.8311	0.8220	0.8126	0.8030
12	0.9313	0.9261	0.9207	0.9150	0.9091	0.9029	0.8965	0.8898	0.8829	0.8758
13	0.9628	0.9595	0.9561	0.9524	0.9486	0.9445	0.9403	0.9358	0.9311	0.9262
14	0.9810	0.9791	0.9771	0.9749	0.9726	0.9701	0.9675	0.9647	0.9617	0.9585
15	0.9908	0.9898	0.9887	0.9875	0.9862	0.9848	0.9832	0.9816	0.9798	0.9780
16	0.9958	0.9953	0.9947	0.9941	0.9934	0.9926	0.9913	0.9909	0.9899	0.9889
17	0.9982	0.9979	0.9977	0.9973	0.9970	0.9966	0.9962	0.9957	0.9952	0.9947
18	0.9992	0.9991	0.9990	0.9989	0.9987	0.9985	0.9983	0.9981	0.9978	0.9976
19	0.9997	0.9997	0.9996	0.9995	0.9995	0.9994	0.9993	0.9992	0.9991	0.9989
20	0.9999	0.9999	0.9998	0.9998	0.9998	0.9998	0.9997	0.9997	0.9996	0.9996
21	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998
22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999
23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

TABLA B (continuación) Valores de la función de distribución acumulativa de Poisson

	λ									
x	9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9	10.0
0	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0000
1	0.0011	0.0010	0.0009	0.0009	0.0008	0.0007	0.0007	0.0006	0.0005	0.0005
2	0.0058	0.0053	0.0049	0.0045	0.0042	0.0038	0.0035	0.0033	0.0030	0.0028
3	0.0198	0.0184	0.0172	0.0160	0.0149	0.0138	0.0129	0.0120	0.0111	0.0103
4	0.0517	0.0486	0.0456	0.0429	0.0403	0.0378	0.0355	0.0333	0.0312	0.0293
5	0.1098	0.1041	0.0987	0.0935	0.0885	0.0838	0.0793	0.0750	0.0710	0.0671
6	0.1978	0.1892	0.1808	0.1727	0.1650	0.1575	0.1502	0.1433	0.1366	0.1301
7	0.3123	0.3010	0.2900	0.2792	0.2687	0.2584	0.2485	0.2388	0.2294	0.2202
8	0.4426	0.4296	0.4168	0.4042	0.3918	0.3796	0.3676	0.3558	0.3442	0.3328
9	0.5742	0.5611	0.5480	0.5349	0.5218	0.5089	0.4960	0.4832	0.4705	0.4579
10	0.6941	0.6820	0.6699	0.6576	0.6453	0.6330	0.6205	0.6081	0.5956	0.5830
11	0.7932	0.7832	0.7730	0.7626	0.7520	0.7412	0.7303	0.7193	0.7081	0.6968
12	0.8684	0.8607	0.8529	0.8448	0.8364	0.8279	0.8191	0.8101	0.8009	0.7916
13	0.9210	0.9156	0.9100	0.9042	0.8981	0.8919	0.8853	0.8786	0.8716	0.8645
14	0.9552	0.9517	0.9480	0.9441	0.9400	0.9357	0.9312	0.9265	0.9216	0.9165
15	0.9760	0.9738	0.9715	0.9691	0.9665	0.9638	0.9609	0.9579	0.9546	0.9513
16	0.9878	0.9865	0.9852	0.9838	0.9823	0.9806	0.9789	0.9770	0.9751	0.9730
17	0.9941	0.9934	0.9927	0.9919	0.9911	0.9902	0.9892	0.9881	0.9870	0.9857
18	0.9973	0.9969	0.9966	0.9962	0.9957	0.9952	0.9947	0.9941	0.9935	0.9928
19	0.9988	0.9986	0.9985	0.9983	0.9980	0.9978	0.9975	0.9972	0.9969	0.9965
20	0.9995	0.9994	0.9993	0.9992	0.9991	0.9990	0.9989	0.9987	0.9986	0.9984
21	0.9998	0.9998	0.9997	0.9997	0.9996	0.9996	0.9995	0.9995	0.9994	0.9993
22	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998	0.9998	0.9997	0.9997
23	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

	λ									
x	11.0	12.0	13.0	14.0	15.0	16.0	17.0	18.0	19.0	20.0
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0012	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0049	0.0023	0.0011	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
4	0.0151	0.0076	0.0037	0.0018	0.0009	0.0004	0.0002	0.0001	0.0000	0.0000
5	0.0375	0.0203	0.0107	0.0055	0.0028	0.0014	0.0007	0.0003	0.0002	0.0001
6	0.0786	0.0458	0.0259	0.0142	0.0076	0.0040	0.0021	0.0010	0.0005	0.0003
7	0.1432	0.0895	0.0540	0.0316	0.0180	0.0100	0.0054	0.0029	0.0015	0.0008
8	0.2320	0.1550	0.0998	0.0621	0.0374	0.0220	0.0126	0.0071	0.0039	0.0021
9	0.3405	0.2424	0.1658	0.1094	0.0699	0.0433	0.0261	0.0154	0.0089	0.0050
10	0.4599	0.3472	0.2517	0.1757	0.1185	0.0774	0.0491	0.0304	0.0183	0.0108
11	0.5793	0.4616	0.3532	0.2600	0.1848	0.1270	0.0847	0.0549	0.0347	0.0214
12	0.6887	0.5760	0.4631	0.3585	0.2676	0.1931	0.1350	0.0917	0.0606	0.0390
13	0.7813	0.6815	0.5730	0.4644	0.3632	0.2745	0.2009	0.1426	0.0984	0.0661
14	0.8540	0.7720	0.6751	0.5704	0.4657	0.3675	0.2808	0.2081	0.1497	0.1049
15	0.9074	0.8444	0.7636	0.6694	0.5681	0.4667	0.3715	0.2867	0.2148	0.1565
16	0.9441	0.8987	0.8355	0.7559	0.6641	0.5660	0.4677	0.3751	0.2920	0.2211
17	0.9678	0.9370	0.8905	0.8272	0.7489	0.6593	0.5640	0.4686	0.3784	0.2970
18	0.9823	0.9626	0.9302	0.8826	0.8195	0.7423	0.6550	0.5622	0.4695	0.3814
19	0.9907	0.9787	0.9573	0.9235	0.8752	0.8122	0.7363	0.6509	0.5606	0.4703
20	0.9953	0.9884	0.9750	0.9521	0.9170	0.8682	0.8055	0.7307	0.6472	0.5591
21	0.9977	0.9939	0.9859	0.9712	0.9469	0.9108	0.8615	0.7991	0.7255	0.6437
22	0.9990	0.9970	0.9924	0.9833	0.9673	0.9418	0.9047	0.8551	0.7931	0.7206
23	0.9995	0.9985	0.9960	0.9907	0.9805	0.9633	0.9367	0.8989	0.8490	0.7875
24	0.9998	0.9993	0.9980	0.9950	0.9888	0.9777	0.9594	0.9317	0.8933	0.8432
25	0.9999	0.9997	0.9990	0.9974	0.9938	0.9869	0.9748	0.9554	0.9269	0.8878
26	1.0000	0.9999	0.9995	0.9987	0.9967	0.9925	0.9848	0.9718	0.9514	0.9221
27	1.0000	0.9999	0.9998	0.9994	0.9983	0.9959	0.9912	0.9827	0.9687	0.9475
28	1.0000	1.0000	0.9999	0.9997	0.9991	0.9978	0.9950	0.9897	0.9805	0.9657
29	1.0000	1.0000	1.0000	0.9999	0.9996	0.9989	0.9973	0.9941	0.9882	0.9782
30	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9986	0.9967	0.9930	0.9865
31	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9993	0.9982	0.9960	0.9919
32	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9990	0.9978	0.9953
33	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9995	0.9988	0.9973
34	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9985
35	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992
36	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996
37	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
38	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999

TABLA C Valores de las funciones de probabilidad y distribución acumulativa para la distribución hipergeométrica

$$P(X \leq x) = F(x; N, n, k) = \sum_{i=0}^{x} \frac{\binom{k}{i}\binom{N-k}{n-i}}{\binom{N}{n}} \qquad p(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

N	n	k	x	$F(x)$	$p(x)$	N	n	k	x	$F(x)$	$p(x)$
2	1	1	0	0.500000	0.500000	6	2	2	2	1.000000	0.066667
2	1	1	1	1.000000	0.500000	6	3	1	0	0.500000	0.500000
3	1	1	0	0.666667	0.666667	6	3	1	1	1.000000	0.500000
3	1	1	1	1.000000	0.333333	6	3	2	0	0.200000	0.200000
3	2	1	0	0.333333	0.333333	6	3	2	1	0.800000	0.600000
3	2	1	1	1.000000	0.666667	6	3	2	2	1,000000	0.200000
3	2	2	1	0.666667	0.666667	6	3	3	0	0.050000	0.050000
3	2	2	2	1.000000	0.333333	6	3	3	1	0.500000	0.450000
4	1	1	0	0.750000	0.750000	6	3	3	2	0.950000	0.450000
4	1	1	1	1.000000	0.250000	6	3	3	3	1.000000	0.050000
4	2	1	0	0.500000	0.500000	6	4	1	0	0.333333	0.333333
4	2	1	1	1.000000	0.500000	6	4	1	1	1.000000	0.666667
4	2	2	0	0.166667	0.166667	6	4	2	0	0.066667	0.066667
4	2	2	1	0.833333	0.666667	6	4	2	1	0.600000	0.533333
4	2	2	2	1.000000	0.166667	6	4	2	2	1.000000	0.400000
4	3	1	0	0.250000	0.250000	6	4	3	1	0.200000	0.200000
4	3	1	1	1.000000	0.750000	6	4	3	2	0.800000	0.600000
4	3	2	1	0.500000	0.500000	6	4	3	3	1.000000	0.200000
4	3	2	2	1.000000	0.500000	6	4	4	2	0.400000	0.400000
4	3	3	2	0.750000	0.750000	6	4	4	3	0.933333	0.533333
4	3	3	3	1.000000	0.250000	6	4	4	4	1.000000	0.066667
5	1	1	0	0.800000	0.800000	6	5	1	0	0.166667	0.166667
5	1	1	1	1.000000	0.200000	6	5	1	1	1.000000	0.833333
5	2	1	0	0.600000	0.600000	6	5	2	1	0.333333	0.333333
5	2	1	1	1.000000	0.400000	6	5	2	2	1.000000	0.666667
5	2	2	0	0.300000	0.300000	6	5	3	2	0.500000	0.500000
5	2	2	1	0.900000	0.600000	6	5	3	3	1.000000	0.500000
5	2	2	2	1.000000	0.100000	6	5	4	3	0.666667	0.666667
5	3	1	0	0.400000	0.400000	6	5	4	4	1.000000	0.333333
5	3	1	1	1.000000	0.600000	6	5	5	4	0.833333	0.833333

5	3	2	0	0.100000	0.100000	6	5	5	5	1.000000	0.166667
5	3	2	1	0.700000	0.600000	7	1	1	0	0.857143	0.857143
5	3	2	2	1.000000	0.300000	7	1	1	1	1.000000	0.142857
5	3	3	1	0.300000	0.300000	7	2	1	0	0.714286	0.714286
5	3	3	2	0.900000	0.600000	7	2	1	1	1.000000	0.285714
5	3	3	3	1.000000	0.100000	7	2	2	0	0.476190	0.476190
5	4	1	0	0.200000	0.200000	7	2	2	1	0.952381	0.476190
5	4	1	1	1.000000	0.800000	7	2	2	2	1.000000	0.047619
5	4	2	1	0.400000	0.400000	7	3	1	0	0.571429	0.571429
5	4	2	2	0.000000	0.600000	7	3	1	1	1.000000	0.428571
5	4	3	2	0.600000	0.600000	7	3	2	0	0.285714	0.285714
5	4	3	3	1.000000	0.400000	7	3	2	1	0.857143	0.571429
5	4	4	3	0.800000	0.800000	7	3	2	2	1.000000	0.142857
5	4	4	4	1.000000	0.200000	7	3	3	0	0.114286	0.114286
6	1	1	0	0.833333	0.833333	7	3	3	1	0.628571	0.514286
6	1	1	1	1.000000	0.166667	7	3	3	2	0.971428	0.342857
6	2	1	0	0.666667	0.666667	7	3	3	3	1.000000	0.028571
6	2	1	1	1.000000	0.333333	7	4	1	0	0.428571	0.428571
6	2	2	0	0.400000	0.400000	7	4	1	1	1.000000	0.571429
6	2	2	1	0.933333	0.533333	7	4	2	0	0.142857	0.142857
7	4	2	1	0.714286	0.571429	8	3	3	2	0.982143	0.267857
7	4	2	2	1.000000	0.285714	8	3	3	3	1.000000	0.017857
7	4	3	0	0.028571	0.028571	8	4	1	0	0.500000	0.500000
7	4	3	1	0.371429	0.342857	8	4	1	1	1.000000	0.500000
7	4	3	2	0.885714	0.514286	8	4	2	0	0.214286	0.214286
7	4	3	3	1.000000	0.114286	8	4	2	1	0.785714	0.571429
7	4	4	1	0.114286	0.114286	8	4	2	2	1.000000	0.214286
7	4	4	2	0.628571	0.514286	8	4	3	0	0.071429	0.071429
7	4	4	3	0.971428	0.342857	8	4	3	1	0.500000	0.428571
7	4	4	4	1.000000	0.028571	8	4	3	2	0.928571	0.428571
7	5	1	0	0.285714	0.285714	8	4	3	3	1.000000	0.071429
7	5	1	1	1.000000	0.714286	8	4	4	0	0.014286	0.014286
7	5	2	0	0.047619	0.047619	8	4	4	1	0.242857	0.228571
7	5	2	1	0.523809	0.476190	8	4	4	2	0.757143	0.514286
7	5	2	2	1.000000	0.476190	8	4	4	3	0.985714	0.228571
7	5	3	1	0.142857	0.142857	8	4	4	4	1.000000	0.014286
7	5	3	2	0.714286	0.571429	8	5	1	0	0.375000	0.375000
7	5	3	3	1.000000	0.285714	8	5	1	1	1.000000	0.625000
7	5	4	2	0.285714	0.285714	8	5	2	0	0.107143	0.107143
7	5	4	3	0.857143	0.571429	8	5	2	1	0.642857	0.535714

TABLA C (continuación) Valores de las funciones de probabilidad y distribución acumulativa para la distribución hipergeométrica

N	n	k	x	$F(x)$	$p(x)$	N	n	k	x	$F(x)$	$p(x)$
7	5	4	4	1.000000	0.142857	8	5	2	2	1.000000	0.357143
7	5	5	3	0.476190	0.476190	8	5	3	0	0.017857	0.017857
7	5	5	4	0.952381	0.476190	8	5	3	1	0.285714	0.267857
7	5	5	5	1.000000	0.047619	8	5	3	2	0.821429	0.535714
7	6	1	0	0.142857	0.142857	8	5	3	3	1.000000	0.178571
7	6	1	1	1.000000	0.857143	8	5	4	1	0.071429	0.071429
7	6	2	1	0.285714	0.285714	8	5	4	2	0.500000	0.428571
7	6	2	2	1.000000	0.714286	8	5	4	3	0.928571	0.428571
7	6	3	2	0.428571	0.428571	8	5	4	4	1.000000	0.071429
7	6	3	3	1.000000	0.571429	8	5	5	2	0.178571	0.178571
7	6	4	3	0.571429	0.571429	8	5	5	3	0.714286	0.535714
7	6	4	4	1.000000	0.428571	8	5	5	4	0.982143	0.267857
7	6	5	4	0.714286	0.714286	8	5	5	5	1.000000	0.017857
7	6	5	5	1.000000	0.285714	8	6	1	0	0.250000	0.250000
7	6	6	5	0.857143	0.857143	8	6	1	1	1.000000	0.750000
7	6	6	6	1.000000	0.142857	8	6	2	0	0.035714	0.035714
8	1	1	0	0.875000	0.875000	8	6	2	1	0.464286	0.428571
8	1	1	1	1.000000	0.125000	8	6	2	2	1.000000	0.535714
8	2	1	0	0.750000	0.750000	8	6	3	1	0.107143	0.107143
8	2	1	1	1.000000	0.250000	8	6	3	2	0.642857	0.535714
8	2	2	0	0.535714	0.535714	8	6	3	3	1.000000	0.357143
8	2	2	1	0.964286	0.428571	8	6	4	2	0.214286	0.214286
8	2	2	2	1.000000	0.035714	8	6	4	3	0.785714	0.571429
8	3	1	0	0.625000	0.625000	8	6	4	4	1.000000	0.214286
8	3	1	1	1.000000	0.375000	8	6	5	3	0.357143	0.357143
8	3	2	0	0.357143	0.357143	8	6	5	4	0.892857	0.535714
8	3	2	1	0.892857	0.535714	8	6	5	5	1.000000	0.107143
8	3	2	2	1.000000	0.107143	8	6	6	4	0.535714	0.535714
8	3	3	0	0.178571	0.178571	8	6	6	5	0.964286	0.428571
8	3	3	1	0.714286	0.535714	8	6	6	6	1.000000	0.035714

8	7	1	0	0.125000	0.125000
8	7	1	1	1.000000	0.875000
8	7	2	1	0.250000	0.250000
8	7	2	2	1.000000	0.750000
8	7	3	2	0.375000	0.375000
8	7	3	3	1.000000	0.625000
8	7	4	3	0.500000	0.500000
8	7	4	4	1.000000	0.500000
8	7	5	4	0.625000	0.625000
8	7	5	5	1.000000	0.375000
8	7	6	5	0.750000	0.750000
8	7	6	6	1.000000	0.250000
8	7	7	6	0.875000	0.875000
8	7	7	7	1.000000	0.125000
9	1	1	0	0.888889	0.888889
9	1	1	1	1.000000	0.111111
9	2	1	0	0.777778	0.777778
9	2	1	1	1.000000	0.222222
9	2	2	0	0.583333	0.583333
9	2	2	1	0.972222	0.388889
9	2	2	2	1.000000	0.027778
9	3	1	0	0.666667	0.666667
9	3	1	1	1.000000	0.333333
9	3	2	0	0.416667	0.416667
9	3	2	1	0.916667	0.500000
9	3	2	2	1.000000	0.083333
9	3	3	0	0.238095	0.238095
9	3	3	1	0.773809	0.535714
9	3	3	2	0.988095	0.214286
9	3	3	3	1.000000	0.011905
9	4	1	0	0.555556	0.555556
9	4	1	1	1.000000	0.444444
9	4	2	0	0.277778	0.277778
9	4	2	1	0.833333	0.555556
9	4	2	2	1.000000	0.166667
9	4	3	0	0.119048	0.119048
9	4	3	1	0.595238	0.476190
9	4	3	2	0.952381	0.357143
9	4	3	3	1.000000	0.047619
9	4	4	0	0.039683	0.039683

9	5	3	1	0.404762	0.357143
9	5	3	2	0.880952	0.476190
9	5	3	3	1.000000	0.119048
9	5	4	0	0.007936	0.007936
9	5	4	1	0.166667	0.158730
9	5	4	2	0.642857	0.476190
9	5	4	3	0.960317	0.317460
9	5	4	4	1.000000	0.039683
9	5	5	1	0.039683	0.039683
9	5	5	2	0.357143	0.317460
9	5	5	3	0.833333	0.476190
9	5	5	4	0.992063	0.158730
9	5	5	5	1.000000	0.007936
9	6	1	0	0.333333	0.333333
9	6	1	1	1.000000	0.666667
9	6	2	0	0.083333	0.083333
9	6	2	1	0.583333	0.500000
9	6	2	2	1.000000	0.416667
9	6	3	0	0.011905	0.011905
9	6	3	1	0.226190	0.214286
9	6	3	2	0.761905	0.535714
9	6	3	3	1.000000	0.238095
9	6	4	1	0.047619	0.047619
9	6	4	2	0.404762	0.357143
9	6	4	3	0.880952	0.476190
9	6	4	4	1.000000	0.119048
9	6	5	2	0.119048	0.119048
9	6	5	3	0.595238	0.476190
9	6	5	4	0.952381	0.357143
9	6	5	5	1.000000	0.047619
9	6	6	3	0.238095	0.238095
9	6	6	4	0.773809	0.535714
9	6	6	5	0.988095	0.214286
9	6	6	6	1.000000	0.011905
9	7	1	0	0.222222	0.222222
9	7	1	1	1.000000	0.777778
9	7	2	0	0.027778	0.027778
9	7	2	1	0.416667	0.388889
9	7	2	2	1.000000	0.583333
9	7	3	1	0.083333	0.083333

TABLA C (continuación) Valores de las funciones de probabilidad y distribución acumulativa para la distribución hipergeométrica

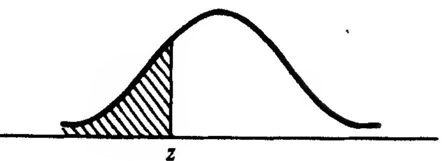
N	n	k	x	$F(x)$	$p(x)$	N	n	k	x	$F(x)$	$p(x)$
9	4	4	1	0.357143	0.317460	9	7	3	2	0.583333	0.500000
9	4	4	2	0.833333	0.476190	9	7	3	3	1.000000	0.416667
9	4	4	3	0.992063	0.158730	9	7	4	2	0.166667	0.166667
9	4	4	4	1.000000	0.007936	9	7	4	3	0.722222	0.555556
9	5	1	0	0.444444	0.444444	9	7	4	4	1.000000	0.277778
9	5	1	1	1.000000	0.555556	9	7	5	3	0.277778	0.277778
9	5	2	0	0.166667	0.166667	9	7	5	4	0.833333	0.555556
9	5	2	1	0.722222	0.555556	9	7	5	5	1.000000	0.166667
9	5	2	2	1.000000	0.277778	9	7	6	4	0.416667	0.416667
9	5	3	0	0.047619	0.047619	9	7	6	5	0.916667	0.500000
9	7	6	6	1.000000	0.833333	10	5	1	0	0.500000	0.500000
9	7	7	5	0.583333	0.583333	10	5	1	1	1.000000	0.500000
9	7	7	6	0.972222	0.388889	10	5	2	0	0.222222	0.222222
9	7	7	7	1.000000	0.027778	10	5	2	1	0.777778	0.555556
9	8	1	0	0.111111	0.111111	10	5	2	2	1.000000	0.222222
9	8	1	1	1.000000	0.888889	10	5	3	0	0.083333	0.083333
9	8	2	1	0.222222	0.222222	10	5	3	1	0.500000	0.416667
9	8	2	2	1.000000	0.777778	10	5	3	2	0.916667	0.416667
9	8	3	2	0.333333	0.333333	10	5	3	3	1.000000	0.083333
9	8	3	3	1.000000	0.666667	10	5	4	0	0.023810	0.023810
9	8	4	3	0.444444	0.444444	10	5	4	1	0.261905	0.238095
9	8	4	4	1.000000	0.555556	10	5	4	2	0.738095	0.476190
9	8	5	4	0.555556	0.555556	10	5	4	3	0.976190	0.238095
9	8	5	5	1.000000	0.444444	10	5	4	4	1.000000	0.023810
9	8	6	5	0.666667	0.666667	10	5	5	0	0.003968	0.003968
9	8	6	6	1.000000	0.333333	10	5	5	1	0.103175	0.099206
9	8	7	6	0.777778	0.777778	10	5	5	2	0.500000	0.396825
9	8	7	7	1.000000	0.222222	10	5	5	3	0.896825	0.396825
9	8	8	7	0.888889	0.888889	10	5	5	4	0.996032	0.099206
9	8	8	8	1.000000	0.111111	10	5	5	5	1.000000	0.003968

10	1	1	0	0.900000	0.900000	10	6	1	0	0.400000	0.400000
10	1	1	1	1.000000	0.100000	10	6	1	1	1.000000	0.600000
10	2	1	0	0.800000	0.800000	10	6	2	0	0.133333	0.133333
10	2	1	1	1.000000	0.200000	10	6	2	1	0.666667	0.533333
10	2	2	0	0.622222	0.622222	10	6	2	2	1.000000	0.333333
10	2	2	1	0.977778	0.355556	10	6	3	0	0.033333	0.033333
10	2	2	2	1.000000	0.022222	10	6	3	1	0.333333	0.300000
10	3	1	0	0.700000	0.700000	10	6	3	2	0.833333	0.500000
10	3	1	1	1.000000	0.300000	10	6	3	3	1.000000	0.166667
10	3	2	0	0.466667	0.466667	10	6	4	0	0.004762	0.004762
10	3	2	1	0.933333	0.466667	10	6	4	1	0.119048	0.114286
10	3	2	2	1.000000	0.066667	10	6	4	2	0.547619	0.428571
10	3	3	0	0.291667	0.291667	10	6	4	3	0.928571	0.380952
10	3	3	1	0.816667	0.525000	10	6	4	4	1.000000	0.071429
10	3	3	2	0.991667	0.175000	10	6	5	1	0.023810	0.023810
10	3	3	3	1.000000	0.008333	10	6	5	2	0.261905	0.238095
10	4	1	0	0.600000	0.600000	10	6	5	3	0.738095	0.476190
10	4	1	1	1.000000	0.400000	10	6	5	4	0.976190	0.238095
10	4	2	0	0.333333	0.333333	10	6	5	5	1.000000	0.023810
10	4	2	1	0.866667	0.533333	10	6	6	2	0.071429	0.071429
10	4	2	2	1.000000	0.133333	10	6	6	3	0.452381	0.380952
10	4	3	0	0.166667	0.166667	10	6	6	4	0.880952	0.428571
10	4	3	1	0.666667	0.500000	10	6	6	5	0.995238	0.114286
10	4	3	2	0.966667	0.300000	10	6	6	6	1.000000	0.004762
10	4	3	3	1.000000	0.033333	10	7	1	0	0.300000	0.300000
10	4	4	0	0.071429	0.071429	10	7	1	1	1.000000	0.700000
10	4	4	1	0.452381	0.380952	10	7	2	0	0.066667	0.066667
10	4	4	2	0.880952	0.428571	10	7	2	1	0.533333	0.466667
10	4	4	3	0.995238	0.114286	10	7	2	2	1.000000	0.466667
10	4	4	4	1.000000	0.004762	10	7	3	0	0.008333	0.008333

Fuente: Tomado de *Tables of the hypergeometric probability distribution*, por Gerald J. Lieberman y Donald B. Owen, con permiso de los editores, Stanford University Press.

TABLA D Valores de la función de distribución acumulativa normal estándar

$$P(Z \leq z) = F(z; 0, 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-t^2/2)dt$$



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183

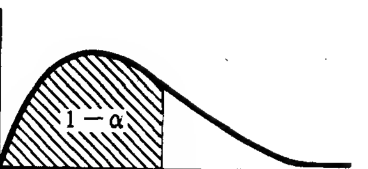
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2297	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389

TABLA D (continuación) Valores de la función de distribución acumulativa normal estándar

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998

TABLA E Valores de cuantiles de la distribución chi-cuadrada

$$P(X \leq x_{1-\alpha,\nu}) = F(x_{1-\alpha};\nu) = \frac{1}{\Gamma(\nu/2)\,2^{\nu/2}}\int_0^{x_{1-\alpha,\nu}} t^{\nu/2-1}\exp(-t/2)dt = 1-\alpha$$



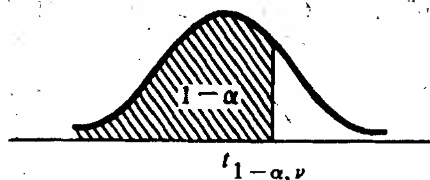
ν	$x_{0.005}$	$x_{0.010}$	$x_{0.025}$	$x_{0.050}$	$x_{0.100}$	$x_{0.900}$	$x_{0.950}$	$x_{0.975}$	$x_{0.990}$	$x_{0.995}$
1	0.00	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.64	7.90
2	0.01	0.02	0.05	0.10	0.21	4.60	5.99	7.38	9.22	10.59
3	0.07	0.11	0.22	0.35	0.58	6.25	7.82	9.36	11.32	12.82
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.15	13.28	14.82
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.84	15.09	16.76
6	0.67	0.87	1.24	1.63	2.20	10.65	12.60	14.46	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.02	18.47	20.27
8	1.34	1.64	2.18	2.73	3.49	13.36	15.51	17.55	20.08	21.94
9	1.73	2.09	2.70	3.32	4.17	14.69	16.93	19.03	21.65	23.56
10	2.15	2.55	3.24	3.94	4.86	15.99	18.31	20.50	23.19	25.15
11	2.60	3.05	3.81	4.57	5.58	17.28	19.68	21.93	24.75	26.71
12	3.06	3.57	4.40	5.22	6.30	18.55	21.03	23.35	26.25	28.25
13	3.56	4.10	5.01	5.89	7.04	19.81	22.37	24.75	27.72	29.88
14	4.07	4.65	5.62	6.57	7.79	21.07	23.69	26.13	29.17	31.38
15	4.59	5.23	6.26	7.26	8.55	22.31	25.00	27.50	30.61	32.86
16	5.14	5.81	6.90	7.96	9.31	23.55	26.30	28.86	32.03	34.32
17	5.69	6.40	7.56	8.67	10.08	24.77	27.59	30.20	33.43	35.77
18	6.25	7.00	8.23	9.39	10.86	25.99	28.88	31.54	34.83	37.21
19	6.82	7.63	8.90	10.11	11.65	27.21	30.15	32.87	36.22	38.63
20	7.42	8.25	9.59	10.85	12.44	28.42	31.42	34.18	37.59	40.05

TABLA E (continuación) Valores de cuantiles de la distribución chi-cuadrada

ν	$x_{0.005}$	$x_{0.010}$	$x_{0.025}$	$x_{0.050}$	$x_{0.100}$	$x_{0.900}$	$x_{0.950}$	$x_{0.975}$	$x_{0.990}$	$x_{0.995}$
21	8.02	8.89	10.28	11.59	13.24	29.62	32.68	35.49	38.96	41.45
22	8.62	9.53	10.98	12.34	14.04	30.82	33.93	36.79	40.31	42.84
23	9.25	10.19	11.69	13.09	14.85	32.01	35.18	38.09	41.66	44.23
24	9.87	10.85	12.40	13.84	15.66	33.20	36.42	39.38	43.00	45.60
25	10.50	11.51	13.11	14.61	16.47	34.38	37.66	40.66	44.34	46.97
26	11.13	12.19	13.84	15.38	17.29	35.57	38.89	41.94	45.66	48.33
27	11.79	12.87	14.57	16.15	18.11	36.74	40.12	43.21	46.99	49.69
28	12.44	13.55	15.30	16.92	18.94	37.92	41.34	44.47	48.30	51.04
29	13.09	14.24	16.04	17.70	19.77	39.09	42.56	45.74	49.61	52.38
30	13.77	14.94	16.78	18.49	20.60	40.26	43.78	46.99	50.91	53.71
35	17.16	18.49	20.56	22.46	24.79	46.06	49.81	53.22	57.36	60.31
40	20.67	22.14	24.42	26.51	29.06	51.80	55.75	59.34	63.71	66.80
45	24.28	25.88	28.36	30.61	33.36	57.50	61.65	65.41	69.98	73.20
50	27.96	29.68	32.35	34.76	37.69	63.16	67.50	71.42	76.17	79.52
60	35.50	37.46	40.47	43.19	46.46	74.39	79.08	83.30	88.40	91.98
70	43.25	45.42	48.75	51.74	55.33	85.52	90.53	95.03	100.44	104.24
80	51.14	53.52	57.15	60.39	64.28	96.57	101.88	106.63	112.34	116.35
90	59.17	61.74	65.64	69.13	73.29	107.56	113.[illegible]4	118.14	124.13	128.32
100	67.30	70.05	74.22	77.93	82.36	118.49	124.34	129.56	135.82	140.19

TABLA F Valores de cuantiles de la distribución t de Student

$$P(T \leq t_{1-\alpha,\nu}) = \frac{\Gamma[(\nu+1)/2]}{\sqrt{\pi\nu}\,\Gamma(\nu/2)} \int_{-\infty}^{t_{1-\alpha,\nu}} [1 + (t^2/\nu)]^{-(\nu+1)/2} dt = 1 - \alpha$$



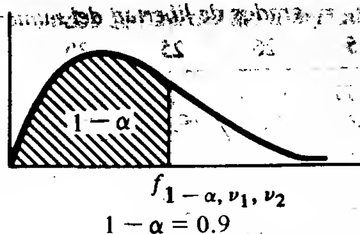
ν	$t_{0.001}$	$t_{0.005}$	$t_{0.010}$	$t_{0.025}$	$t_{0.050}$	$t_{0.100}$	$t_{0.200}$
1	−318.309	−63.657	−31.821	−12.706	−6.314	−3.078	−1.376
2	−22.327	−9.925	−6.965	−4.303	−2.920	−1.886	−1.061
3	−10.215	−5.841	−4.541	−3.182	−2.353	−1.638	−0.978
4	−7.173	−4.604	−3.747	−2.776	−2.132	−1.533	−0.941
5	−5.893	−4.032	−3.365	−2.571	−2.015	−1.476	−0.920
6	−5.208	−3.707	−3.143	−2.447	−1.943	−1.440	−0.906
7	−4.785	−3.499	−2.998	−2.365	−1.895	−1.415	−0.896
8	−4.501	−3.355	−2.896	−2.306	−1.860	−1.397	−0.889
9	−4.297	−3.250	−2.821	−2.262	−1.833	−1.383	−0.883
10	−4.144	−3.169	−2.764	−2.228	−1.812	−1.372	−0.879
11	−4.025	−3.106	−2.718	−2.201	−1.796	−1.363	−0.876
12	−3.930	−3.055	−2.681	−2.179	−1.782	−1.356	−0.873
13	−3.852	−3.012	−2.650	−2.160	−1.771	−1.350	−0.870
14	−3.787	−2.977	−2.624	−2.145	−1.761	−1.345	−0.868
15	−3.733	−2.947	−2.602	−2.131	−1.753	−1.341	−0.866
16	−3.686	−2.921	−2.583	−2.120	−1.746	−1.337	−0.865
17	−3.646	−2.898	−2.567	−2.110	−1.740	−1.333	−0.863
18	−3.610	−2.878	−2.552	−2.101	−1.734	−1.330	−0.862
19	−3.579	−2.861	−2.539	−2.093	−1.729	−1.328	−0.861
20	−3.552	−2.845	−2.528	−2.086	−1.725	−1.325	−0.860
21	−3.527	−2.831	−2.518	−2.080	−1.721	−1.323	−0.859
22	−3.505	−2.819	−2.508	−2.074	−1.717	−1.321	−0.858
23	−3.485	−2.807	−2.500	−2.069	−1.714	−1.319	−0.858
24	−3.467	−2.797	−2.492	−2.064	−1.711	−1.318	−0.857
25	−3.450	−2.787	−2.485	−2.060	−1.708	−1.316	−0.856
26	−3.435	−2.779	−2.479	−2.056	−1.706	−1.315	−0.856
27	−3.421	−2.771	−2.473	−2.052	−1.703	−1.314	−0.855
28	−3.408	−2.763	−2.467	−2.048	−1.701	−1.313	−0.855
29	−3.396	−2.756	−2.462	−2.045	−1.699	−1.311	−0.854
30	−3.385	−2.750	−2.457	−2.042	−1.697	−1.310	−0.854
35	−3.340	−2.724	−2.438	−2.030	−1.690	−1.306	−0.852
40	−3.307	−2.704	−2.423	−2.021	−1.684	−1.303	−0.851
45	−3.281	−2.690	−2.412	−2.014	−1.679	−1.301	−0.850
50	−3.261	−2.678	−2.403	−2.009	−1.676	−1.299	−0.849
60	−3.232	−2.660	−2.390	−2.000	−1.671	−1.296	−0.848
70	−3.211	−2.648	−2.381	−1.994	−1.667	−1.294	−0.847
80	−3.195	−2.639	−2.374	−1.990	−1.664	−1.292	−0.846
90	−3.183	−2.632	−2.369	−1.987	−1.662	−1.291	−0.846
100	−3.174	−2.626	−2.364	−1.984	−1.660	−1.290	−0.845
200	−3.131	−2.601	−2.345	−1.972	−1.652	−1.286	−0.843
500	−3.107	−2.586	−2.334	−1.965	−1.648	−1.283	−0.842
1000	−3.098	−2.581	−2.330	−1.962	−1.646	−1.282	−0.842

TABLA F (continuación) Valores de cuantiles de la distribución *t* de Student

ν	$t_{0.800}$	$t_{0.900}$	$t_{0.950}$	$t_{0.975}$	$t_{0.990}$	$t_{0.995}$	$t_{0.999}$
1	1.376	3.078	6.314	12.706	31.820	63.656	318.294
2	1.061	1.886	2.920	4.303	6.965	9.925	22.327
3	0.978	1.638	2.353	3.182	4.541	5.841	10.214
4	0.941	1.533	2.132	2.776	3.747	4.604	7.173
5	0.920	1.476	2.015	2.571	3.365	4.032	5.893
6	0.906	1.440	1.943	2.447	3.143	3.707	5.208
7	0.896	1.415	1.895	2.365	2.998	3.499	4.785
8	0.889	1.397	1.860	2.306	2.896	3.355	4.501
9	0.883	1.383	1.833	2.262	2.821	3.250	4.297
10	0.879	1.372	1.812	2.228	2.764	3.169	4.144
11	0.876	1.363	1.796	2.201	2.718	3.106	4.025
12	0.873	1.356	1.782	2.179	2.681	3.055	3.930
13	0.870	1.350	1.771	2.160	2.650	3.012	3.852
14	0.868	1.345	1.761	2.145	2.624	2.977	3.787
15	0.866	1.341	1.753	2.131	2.602	2.947	3.733
16	0.865	1.337	1.746	2.120	2.583	2.921	3.686
17	0.863	1.333	1.740	2.110	2.567	2.898	3.646
18	0.862	1.330	1.734	2.101	2.552	2.878	3.610
19	0.861	1.328	1.729	2.093	2.539	2.861	3.579
20	0.860	1.325	1.725	2.086	2.528	2.845	3.552
21	0.859	1.323	1.721	2.080	2.518	2.831	3.527
22	0.858	1.321	1.717	2.074	2.508	2.819	3.505
23	0.858	1.319	1.714	2.069	2.500	2.807	3.485
24	0.857	1.318	1.711	2.064	2.492	2.797	3.467
25	0.856	1.316	1.708	2.060	2.485	2.787	3.450
26	0.856	1.315	1.706	2.056	2.479	2.779	3.435
27	0.855	1.314	1.703	2.052	2.473	2.771	3.421
28	0.855	1.313	1.701	2.048	2.467	2.763	3.408
29	0.854	1.311	1.699	2.045	2.462	2.756	3.396
30	0.854	1.310	1.697	2.042	2.457	2.750	3.385
35	0.852	1.306	1.690	2.030	2.438	2.724	3.340
40	0.851	1.303	1.684	2.021	2.423	2.704	3.307
45	0.850	1.301	1.679	2.014	2.412	2.690	3.281
50	0.849	1.299	1.676	2.009	2.403	2.678	3.261
60	0.848	1.296	1.671	2.000	2.390	2.660	3.232
70	0.847	1.294	1.667	1.994	2.381	2.648	3.211
80	0.846	1.292	1.664	1.990	2.374	2.639	3.195
90	0.846	1.291	1.662	1.987	2.368	2.632	3.183
100	0.845	1.290	1.660	1.984	2.364	2.626	3.174
200	0.843	1.286	1.652	1.972	2.345	2.601	3.131
500	0.842	1.283	1.648	1.965	2.334	2.586	3.107
1000	0.842	1.282	1.646	1.962	2.330	2.581	3.098

TABLA G Valores de cuantiles de la distribución F

$$P(F \leq f_{1-\alpha,\nu_1,\nu_2}) = \frac{\Gamma[(\nu_1+\nu_2)/2]\nu_1^{\nu_1/2}\nu_2^{\nu_2/2}}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \int_0^{f_{1-\alpha,\nu_1,\nu_2}} t^{(\nu_1-2)/2}(\nu_2+\nu_1 t)^{-(\nu_1+\nu_2)/2}\,dt = 1-\alpha$$



$1 - \alpha = 0.9$

ν_2	ν_1 = grados de libertad del numerador: 1	2	3	4	5	6	7	8	9	10
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94
7	3.59	3.26	3.07	2.96	2.88	2.83	2.79	2.75	2.72	2.70
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82
35	2.85	2.46	2.25	2.11	2.02	1.95	1.90	1.85	1.82	1.79
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76
50	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.76	1.73
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71
80	2.77	2.37	2.15	2.02	1.92	1.85	1.79	1.75	1.71	1.68
100	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.69	1.66
200	2.73	2.33	2.11	1.97	1.88	1.80	1.75	1.70	1.66	1.63
500	2.72	2.31	2.09	1.96	1.86	1.79	1.73	1.68	1.64	1.61
1000	2.71	2.31	2.09	1.95	1.85	1.78	1.72	1.68	1.64	1.61

TABLA G (continuación) Valores de cuantiles de la distribución F

$1-\alpha = 0.9$

ν_1 = grados de libertad del numerador

ν_2	11	12	15	20	25	30	40	50	100	1000
1	60.47	60.71	61.22	61.74	62.06	62.26	62.53	62.69	63.00	63.29
2	9.40	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.22	5.22	5.20	5.19	5.17	5.17	5.16	5.15	5.14	5.13
4	3.91	3.90	3.87	3.84	3.83	3.82	3.80	3.80	3.78	3.76
5	3.28	3.27	3.24	3.21	3.19	3.17	3.16	3.15	3.13	3.11
6	2.92	2.90	2.87	2.84	2.81	2.80	2.78	2.77	2.75	2.72
7	2.68	2.67	2.63	2.59	2.57	2.56	2.54	2.52	2.50	2.47
8	2.52	2.50	2.46	2.42	2.40	2.38	2.36	2.35	2.32	2.30
9	2.40	2.38	2.34	2.30	2.27	2.25	2.23	2.22	2.19	2.16
10	2.30	2.28	2.24	2.20	2.17	2.16	2.13	2.12	2.09	2.06
11	2.23	2.21	2.17	2.12	2.10	2.08	2.05	2.04	2.00	1.98
12	2.17	2.15	2.10	2.06	2.03	2.01	1.99	1.97	1.94	1.91
13	2.12	2.10	2.05	2.01	1.98	1.96	1.93	1.92	1.88	1.85
14	2.07	2.05	2.01	1.96	1.93	1.91	1.89	1.87	1.83	1.80
15	2.04	2.02	1.97	1.92	1.89	1.87	1.85	1.83	1.79	1.76
16	2.01	1.99	1.94	1.89	1.86	1.84	1.81	1.79	1.76	1.72
17	1.98	1.96	1.91	1.86	1.83	1.81	1.78	1.76	1.73	1.69
18	1.95	1.93	1.89	1.84	1.80	1.78	1.75	1.74	1.70	1.66
19	1.93	1.91	1.86	1.81	1.78	1.76	1.73	1.71	1.67	1.64
20	1.91	1.89	1.84	1.79	1.76	1.74	1.71	1.69	1.65	1.61
21	1.90	1.87	1.83	1.78	1.74	1.72	1.69	1.67	1.63	1.59
22	1.88	1.86	1.81	1.76	1.73	1.70	1.67	1.65	1.61	1.57
23	1.87	1.84	1.80	1.74	1.71	1.69	1.66	1.64	1.59	1.55
24	1.85	1.83	1.78	1.73	1.70	1.67	1.64	1.62	1.58	1.54
25	1.84	1.82	1.77	1.72	1.68	1.66	1.63	1.61	1.56	1.52
26	1.83	1.81	1.76	1.71	1.67	1.65	1.61	1.59	1.55	1.51
27	1.82	1.80	1.75	1.70	1.66	1.64	1.60	1.58	1.54	1.50
28	1.81	1.79	1.74	1.69	1.65	1.63	1.59	1.57	1.53	1.48
29	1.80	1.78	1.73	1.68	1.64	1.62	1.58	1.56	1.52	1.47
30	1.79	1.77	1.72	1.67	1.63	1.61	1.57	1.55	1.51	1.46
35	1.76	1.74	1.69	1.63	1.60	1.57	1.53	1.51	1.47	1.42
40	1.74	1.71	1.66	1.61	1.57	1.54	1.51	1.48	1.43	1.38
50	1.70	1.68	1.63	1.57	1.53	1.50	1.46	1.44	1.39	1.33
60	1.68	1.66	1.60	1.54	1.50	1.48	1.44	1.41	1.36	1.30
80	1.65	1.63	1.57	1.51	1.47	1.44	1.40	1.38	1.32	1.25
100	1.64	1.61	1.56	1.49	1.45	1.42	1.38	1.35	1.29	1.22
200	1.60	1.58	1.52	1.46	1.41	1.38	1.34	1.31	1.24	1.16
500	1.58	1.56	1.50	1.44	1.39	1.36	1.31	1.28	1.21	1.11
1000	1.58	1.55	1.49	1.43	1.38	1.35	1.30	1.27	1.20	1.08

TABLA G (continuación) Valores de cuantiles de la distribución F

$1 - \alpha = 0.95$

ν_2	ν_1 = *grados de libertad del numerador* 1	2	3	4	5	6	7	8	9	10
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.97
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.73
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85
1000	3.85	3.01	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84

TABLA G (continuación) Valores de cuantiles de la distribución F

$1 - \alpha = 0.95$

	ν_1 = *grados de libertad del numerador*									
ν_2	11	12	15	20	25	30	40	50	100	1000
1	242.98	243.91	245.96	248.01	249.26	250.08	251.15	251.77	253.01	254.17
2	19.40	19.41	19.43	19.45	19.46	19.46	19.47	19.48	19.49	19.50
3	8.76	8.74	8.70	8.66	8.63	8.62	8.59	8.58	8.55	8.53
4	5.94	5.91	5.86	5.80	5.77	5.74	5.72	5.70	5.66	5.63
5	4.70	4.68	4.62	4.56	4.52	4.50	4.46	4.44	4.41	4.37
6	4.03	4.00	3.94	3.87	3.84	3.81	3.77	3.75	3.71	3.67
7	3.60	3.57	3.51	3.44	3.40	3.38	3.34	3.32	3.27	3.23
8	3.31	3.28	3.22	3.15	3.11	3.08	3.04	3.02	2.97	2.93
9	3.10	3.07	3.01	2.94	2.89	2.86	2.83	2.80	2.76	2.71
10	2.94	2.91	2.85	2.77	2.73	2.70	2.66	2.64	2.59	2.54
11	2.82	2.79	2.72	2.65	2.60	2.57	2.53	2.51	2.46	2.41
12	2.72	2.69	2.62	2.54	2.50	2.47	2.43	2.40	2.35	2.30
13	2.63	2.60	2.53	2.46	2.41	2.38	2.34	2.31	2.26	2.21
14	2.57	2.53	2.46	2.39	2.34	2.31	2.27	2.24	2.19	2.14
15	2.51	2.48	2.40	2.33	2.28	2.25	2.20	2.18	2.12	2.07
16	2.46	2.42	2.35	2.28	2.23	2.19	2.15	2.12	2.07	2.02
17	2.41	2.38	2.31	2.23	2.18	2.15	2.10	2.08	2.02	1.97
18	2.37	2.34	2.27	2.19	2.14	2.11	2.06	2.04	1.98	1.92
19	2.34	2.31	2.23	2.16	2.11	2.07	2.03	2.00	1.94	1.88
20	2.31	2.28	2.20	2.12	2.07	2.04	1.99	1.97	1.91	1.85
21	2.28	2.25	2.18	2.10	2.05	2.01	1.96	1.94	1.88	1.82
22	2.26	2.23	2.15	2.07	2.02	1.98	1.94	1.91	1.85	1.79
23	2.24	2.20	2.13	2.05	2.00	1.96	1.91	1.88	1.82	1.76
24	2.22	2.18	2.11	2.03	1.97	1.94	1.89	1.86	1.80	1.74
25	2.20	2.16	2.09	2.01	1.96	1.92	1.87	1.84	1.78	1.72
26	2.18	2.15	2.07	1.99	1.94	1.90	1.85	1.82	1.76	1.70
27	2.17	2.13	2.06	1.97	1.92	1.88	1.84	1.81	1.74	1.68
28	2.15	2.12	2.04	1.96	1.91	1.87	1.82	1.79	1.73	1.66
29	2.14	2.10	2.03	1.94	1.89	1.85	1.81	1.77	1.71	1.65
30	2.13	2.09	2.01	1.93	1.88	1.84	1.79	1.76	1.70	1.63
35	2.07	2.04	1.96	1.88	1.82	1.79	1.74	1.70	1.63	1.57
40	2.04	2.00	1.92	1.84	1.78	1.74	1.69	1.66	1.59	1.52
50	1.99	1.95	1.87	1.78	1.73	1.69	1.63	1.60	1.52	1.45
60	1.95	1.92	1.84	1.75	1.69	1.65	1.59	1.56	1.48	1.40
80	1.91	1.88	1.79	1.70	1.64	1.60	1.54	1.51	1.43	1.34
100	1.89	1.85	1.77	1.68	1.62	1.57	1.52	1.48	1.39	1.30
200	1.84	1.80	1.72	1.62	1.56	1.52	1.46	1.41	1.32	1.21
500	1.81	1.77	1.69	1.59	1.53	1.48	1.42	1.38	1.28	1.14
1000	1.80	1.76	1.68	1.58	1.52	1.47	1.41	1.36	1.26	1.11

TABLA G (continuación) Valores de cuantiles de la distribución *F*

$1-\alpha = 0.99$

	ν_1 = *grados de libertad del numerador*									
ν_2	1	2	3	4	5	6	7	8	9	10
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.50	27.34	27.22
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
35	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50
200	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41
500	6.69	4.65	3.82	3.36	3.05	2.84	2.68	2.55	2.44	2.36
1000	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34

TABLA G (continuación) Valores de cuantiles de la distribución F

$1 - \alpha = 0.99$

	ν_1 = *grados de libertad del numerador*									
ν_2	11	12	15	20	25	30	40	50	100	1000
2	99.41	99.42	99.43	99.45	99.46	99.46	99.47	99.48	99.49	99.51
3	27.12	27.03	26.85	26.67	26.58	26.50	26.41	26.35	26.24	26.14
4	14.45	14.37	14.19	14.02	13.91	13.84	13.75	13.69	13.58	13.48
5	9.96	9.89	9.72	9.55	9.45	9.38	9.30	9.24	9.13	9.03
6	7.79	7.72	7.56	7.40	7.29	7.23	7.15	7.09	6.99	6.89
7	6.54	6.47	6.31	6.16	6.06	5.99	5.91	5.86	5.75	5.66
8	5.73	5.67	5.52	5.36	5.26	5.20	5.12	5.07	4.96	4.87
9	5.18	5.11	4.96	4.81	4.71	4.65	4.57	4.52	4.41	4.32
10	4.77	4.71	4.56	4.41	4.31	4.25	4.17	4.12	4.01	3.92
11	4.46	4.40	4.25	4.10	4.00	3.94	3.86	3.81	3.71	3.61
12	4.22	4.16	4.01	3.86	3.76	3.70	3.62	3.57	3.47	3.37
13	4.02	3.96	3.82	3.66	3.57	3.51	3.43	3.38	3.27	3.18
14	3.86	3.80	3.66	3.51	3.41	3.35	3.27	3.22	3.11	3.02
15	3.73	3.67	3.52	3.37	3.28	3.21	3.13	3.08	2.98	2.88
16	3.62	3.55	3.41	3.26	3.16	3.10	3.02	2.97	2.86	2.76
17	3.52	3.46	3.31	3.16	3.07	3.00	2.92	2.87	2.76	2.66
18	3.43	3.37	3.23	3.08	2.98	2.92	2.84	2.78	2.68	2.58
19	3.36	3.30	3.15	3.00	2.91	2.84	2.76	2.71	2.60	2.50
20	3.29	3.23	3.09	2.94	2.84	2.78	2.69	2.64	2.54	2.43
21	3.24	3.17	3.03	2.88	2.78	2.72	2.64	2.58	2.48	2.37
22	3.18	3.12	2.98	2.83	2.73	2.67	2.58	2.53	2.42	2.32
23	3.14	3.07	2.93	2.78	2.69	2.62	2.54	2.48	2.37	2.27
24	3.09	3.03	2.89	2.74	2.64	2.58	2.49	2.44	2.33	2.22
25	3.06	2.99	2.85	2.70	2.60	2.54	2.45	2.40	2.29	2.18
26	3.02	2.96	2.81	2.66	2.57	2.50	2.42	2.36	2.25	2.14
27	2.99	2.93	2.78	2.63	2.54	2.47	2.38	2.33	2.22	2.11
28	2.96	2.90	2.75	2.60	2.51	2.44	2.35	2.30	2.19	2.08
29	2.93	2.87	2.73	2.57	2.48	2.41	2.33	2.27	2.16	2.05
30	2.91	2.84	2.70	2.55	2.45	2.39	2.30	2.24	2.13	2.02
35	2.80	2.74	2.60	2.44	2.35	2.28	2.19	2.14	2.02	1.90
40	2.73	2.66	2.52	2.37	2.27	2.20	2.11	2.06	1.94	1.82
50	2.62	2.56	2.42	2.27	2.17	2.10	2.01	1.95	1.82	1.70
60	2.56	2.50	2.35	2.20	2.10	2.03	1.94	1.88	1.75	1.62
80	2.48	2.42	2.27	2.12	2.01	1.94	1.85	1.79	1.65	1.51
100	2.43	2.37	2.22	2.07	1.97	1.89	1.80	1.74	1.60	1.45
200	2.34	2.27	2.13	1.97	1.87	1.79	1.69	1.63	1.48	1.30
500	2.28	2.22	2.07	1.92	1.81	1.74	1.63	1.57	1.41	1.20
1000	2.27	2.20	2.06	1.90	1.79	1.72	1.61	1.54	1.38	1.16

TABLA H *k*-valores para los límites de tolerancia bilaterales cuando se muestrean distribuciones normales

$n \backslash d$	$\gamma = 0.75$ 0.75	0.90	0.95	0.99	0.999	$\gamma = 0.90$ 0.75	0.90	0.95	0.99	0.999
6	1.704	2.429	2.889	3.779	4.802	2.196	3.131	3.723	4.870	6.188
7	1.624	2.318	2.757	3.611	4.593	2.034	2.902	3.452	4.521	5.750
8	1.568	2.238	2.663	3.491	4.444	1.921	2.743	3.264	4.278	5.446
9	1.525	2.178	2.593	3.400	4.330	1.839	2.626	3.125	4.098	5.220
10	1.492	2.131	2.537	3.328	4.241	1.775	2.535	3.018	3.959	5.046
11	1.465	2.093	2.493	3.271	4.169	1.724	2.463	2.933	3.849	4.906
12	1.443	2.062	2.456	3.223	4.110	1.683	2.404	2.863	3.758	4.792
13	1.425	2.036	2.424	3.183	4.059	1.648	2.355	2.805	3.682	4.697
14	1.409	2.013	2.398	3.148	4.016	1.619	2.314	2.756	3.618	4.615
15	1.395	1.994	2.375	3.118	3.979	1.594	2.278	2.713	3.562	4.545
16	1.383	1.977	2.355	3.092	3.946	1.572	2.246	2.676	3.514	4.484
17	1.372	1.962	2.337	3.069	3.917	1.552	2.219	2.643	3.471	4.430
18	1.363	1.948	2.321	3.048	3.891	1.535	2.194	2.614	3.433	4.382
19	1.355	1.936	2.307	3.030	3.867	1.520	2.172	2.588	3.399	4.339
20	1.347	1.925	2.294	3.013	3.846	1.506	2.152	2.564	3.368	4.300
21	1.340	1.915	2.282	2.998	3.827	1.493	2.135	2.543	3.340	4.264
22	1.334	1.906	2.271	2.984	3.809	1.482	2.118	2.524	3.315	4.232
23	1.328	1.898	2.261	2.971	3.793	1.471	2.103	2.506	3.292	4.203
24	1.322	1.891	2.252	2.959	3.778	1.462	2.089	2.489	3.270	4.176
25	1.317	1.883	2.244	2.948	3.764	1.453	2.077	2.474	3.251	4.151
26	1.313	1.877	2.236	2.938	3.751	1.444	2.065	2.460	3.232	4.127
27	1.309	1.871	2.229	2.929	3.740	1.437	2.054	2.447	3.215	4.106
28	1.305	1.865	2.222	2.920	3.728	1.430	2.044	2.435	3.199	4.085
29	1.301	1.860	2.216	2.911	3.718	1.423	2.034	2.424	3.184	4.066
30	1.297	1.855	2.210	2.904	3.708	1.417	2.025	2.413	3.170	4.049
31	1.294	1.850	2.204	2.896	3.699	1.411	2.017	2.403	3.157	4.032
32	1.291	1.846	2.199	2.890	3.690	1.405	2.009	2.393	3.145	4.016
33	1.288	1.842	2.194	2.883	3.682	1.400	2.001	2.385	3.133	4.001
34	1.285	1.838	2.189	2.877	3.674	1.395	1.994	2.376	3.122	3.987
35	1.283	1.834	2.185	2.871	3.667	1.390	1.988	2.368	3.112	3.974
36	1.280	1.830	2.181	2.866	3.660	1.386	1.981	2.361	3.102	3.961
37	1.278	1.827	2.177	2.860	3.653	1.381	1.975	2.353	3.092	3.949
38	1.275	1.824	2.173	2.855	3.647	1.377	1.969	2.346	3.083	3.938
39	1.273	1.821	2.169	2.850	3.641	1.374	1.964	2.340	3.075	3.927
40	1.271	1.818	2.166	2.846	3.635	1.370	1.959	2.334	3.066	3.917
41	1.269	1.815	2.162	2.841	3.629	1.366	1.954	2.328	3.059	3.907
42	1.267	1.812	2.159	2.837	3.624	1.363	1.949	2.322	3.051	3.897
43	1.266	1.810	2.156	2.833	3.619	1.360	1.944	2.316	3.044	3.888
44	1.264	1.807	2.153	2.829	3.614	1.357	1.940	2.311	3.037	3.879
45	1.262	1.805	2.150	2.826	3.609	1.354	1.935	2.306	3.030	3.871
46	1.261	1.802	2.148	2.822	3.605	1.351	1.931	2.301	3.024	3.863
47	1.259	1.800	2.145	2.819	3.600	1.348	1.927	2.297	3.018	3.855
48	1.258	1.798	2.143	2.815	3.596	1.345	1.924	2.292	3.012	3.847
49	1.256	1.796	2.140	2.812	3.592	1.343	1.920	2.288	3.006	3.840
50	1.255	1.794	2.138	2.809	3.588	1.340	1.916	2.284	3.001	3.833

TABLA H (continuación) *k*-valores para los límites de tolerancia bilaterales cuando se muestrean distribuciones normales

$n \backslash d$	$\gamma = 0.95$					$\gamma = 0.99$				
	0.75	0.90	0.95	0.99	0.999	0.75	0.90	0.95	0.99	0.999
6	2.604	3.712	4.414	5.775	7.337	3.743	5.337	6.345	8.301	10.548
7	2.361	3.369	4.007	5.248	6.676	3.233	4.613	5.488	7.187	9.142
8	2.197	3.136	3.732	4.891	6.226	2.905	4.147	4.936	6.468	8.234
9	2.078	2.967	3.532	4.631	5.899	2.677	3.822	4.550	5.966	7.600
10	1.987	2.839	3.379	4.433	5.649	2.508	3.582	4.265	5.594	7.129
11	1.916	2.737	3.259	4.277	5.452	2.378	3.397	4.045	5.308	6.766
12	1.858	2.655	3.162	4.150	5.291	2.274	3.250	3.870	5.079	6.477
13	1.810	2.587	3.081	4.044	5.158	2.190	3.130	3.727	4.893	6.240
14	1.770	2.529	3.012	3.955	5.045	2.120	3.029	3.608	4.737	6.043
15	1.735	2.480	2.954	3.878	4.949	2.060	2.945	3.507	4.605	5.876
16	1.705	2.437	2.903	3.812	4.865	2.009	2.872	3.421	4.492	5.732
17	1.679	2.400	2.858	3.754	4.791	1.965	2.808	3.345	4.393	5.607
18	1.655	2.366	2.819	3.702	4.725	1.926	2.753	3.279	4.307	5.497
19	1.635	2.337	2.784	3.656	4.667	1.891	2.703	3.221	4.230	5.399
20	1.616	2.310	2.752	3.615	4.614	1.860	2.659	3.168	4.161	5.312
21	1.599	2.286	2.723	3.577	4.567	1.833	2.620	3.121	4.100	5.234
22	1.584	2.264	2.697	3.543	4.523	1.808	2.584	3.078	4.044	5.163
23	1.570	2.244	2.673	3.512	4.484	1.785	2.551	3.040	3.993	5.098
24	1.557	2.225	2.651	3.483	4.447	1.764	2.522	3.004	3.947	5.039
25	1.545	2.208	2.631	3.457	4.413	1.745	2.494	2.972	3.904	4.985
26	1.534	2.193	2.612	3.432	4.382	1.727	2.469	2.941	3.865	4.935
27	1.523	2.178	2.595	3.409	4.353	1.711	2.446	2.914	3.828	4.888
28	1.514	2.164	2.579	3.388	4.326	1.695	2.424	2.888	3.794	4.845
29	1.505	2.152	2.554	3.368	4.301	1.681	2.404	2.864	3.763	4.805
30	1.497	2.140	2.549	3.350	4.278	1.668	2.385	2.841	3.733	4.768
31	1.489	2.129	2.536	3.332	4.256	1.656	2.367	2.820	3.706	4.732
32	1.481	2.118	2.524	3.316	4.235	1.644	2.351	2.801	3.680	4.699
33	1.475	2.108	2.512	3.300	4.215	1.633	2.335	2.782	3.655	4.668
34	1.468	2.099	2.501	3.286	4.197	1.623	2.320	2.764	3.632	4.639
35	1.462	2.090	2.490	3.272	4.179	1.613	2.306	2.748	3.611	4.611
36	1.455	2.081	2.479	3.258	4.161	1.604	2.293	2.732	3.590	4.585
37	1.450	2.073	2.470	3.246	4.146	1.595	2.281	2.717	3.571	4.560
38	1.446	2.068	2.464	3.237	4.134	1.587	2.269	2.703	3.552	4.537
39	1.441	2.060	2.455	3.226	4.120	1.579	2.257	2.690	3.534	4.514
40	1.435	2.052	2.445	3.213	4.104	1.571	2.247	2.677	3.518	4.493
41	1.430	2.045	2.437	3.202	4.090	1.564	2.236	2.665	3.502	4.472
42	1.426	2.039	2.429	3.192	4.077	1.557	2.227	2.653	3.486	4.453
43	1.422	2.033	2.422	3.183	4.065	1.551	2.217	2.642	3.472	4.434
44	1.418	2.027	2.415	3.173	4.053	1.545	2.208	2.631	3.458	4.416
45	1.414	2.021	2.408	3.165	4.042	1.539	2.200	2.621	3.444	4.399
46	1.410	2.016	2.402	3.156	4.031	1.533	2.192	2.611	3.431	4.383
47	1.406	2.011	2.396	3.148	4.021	1.527	2.184	2.602	3.419	4.367
48	1.403	2.006	2.390	3.140	4.011	1.522	2.176	2.593	3.407	4.352
49	1.399	2.001	2.384	3.133	4.002	1.517	2.169	2.584	3.396	4.337
50	1.396	1.969	2.379	3.126	3.993	1.512	2.162	2.576	3.385	4.323

Source: C. Eisenhart, M. W. Hastay, and W. A. Wallis, *Techniques of statistical analysis,* McGraw-Hill, New York, 1947. Publicado con permiso.

TABLA I k-valores para los límites de tolerancia unilaterales cuando se muestrean distribuciones normales

n \ d	$\gamma = 0.75$					$\gamma = 0.90$				
	0.75	0.90	0.95	0.99	0.999	0.75	0.90	0.95	0.99	0.999
6	1.087	1.860	2.336	3.243	4.273	1.540	2.494	3.091	4.242	5.556
7	1.043	1.791	2.250	3.126	4.118	1.435	2.333	2.894	3.972	5.201
8	1.010	1.740	2.190	3.042	4.008	1.360	2.219	2.755	3.783	4.955
9	0.984	1.702	2.141	2.977	3.924	1.302	2.133	2.649	3.641	4.772
10	0.964	1.671	2.103	2.927	3.858	1.257	2.065	2.568	3.532	4.629
11	0.947	1.646	2.073	2.885	3.804	1.219	2.012	2.503	3.444	4.515
12	0.933	1.624	2.048	2.851	3.760	1.188	1.966	2.448	3.371	4.420
13	0.919	1.606	2.026	2.822	3.722	1.162	1.928	2.403	3.310	4.341
14	0.909	1.591	2.007	2.796	3.690	1.139	1.895	2.363	3.257	4.274
15	0.899	1.577	1.991	2.776	3.661	1.119	1.866	2.329	3.212	4.215
16	0.891	1.566	1.977	2.756	3.637	1.101	1.842	2.299	3.172	4.164
17	0.883	1.554	1.964	2.739	3.615	1.085	1.820	2.272	3.136	4.118
18	0.876	1.544	1.951	2.723	3.595	1.071	1.800	2.249	3.106	4.078
19	0.870	1.536	1.942	2.710	3.577	1.058	1.781	2.228	3.078	4.041
20	0.865	1.528	1.933	2.697	3.561	1.046	1.765	2.208	3.052	4.009
21	0.859	1.520	1.923	2.686	3.545	1.035	1.750	2.190	3.028	3.979
22	0.854	1.514	1.916	2.675	3.532	1.025	1.736	2.174	3.007	3.952
23	0.849	1.508	1.907	2.665	3.520	1.016	1.724	2.159	2.987	3.927
24	0.845	1.502	1.901	2.656	3.509	1.007	1.712	2.145	2.969	3.904
25	0.842	1.496	1.895	2.647	3.497	0.999	1.702	2.132	2.952	3.882
30	0.825	1.475	1.869	2.613	3.454	0.966	1.657	2.080	2.884	3.794
35	0.812	1.458	1.849	2.588	3.421	0.942	1.623	2.041	2.833	3.730
40	0.803	1.445	1.834	2.568	3.395	0.923	1.598	2.010	2.793	3.679
45	0.795	1.435	1.821	2.552	3.375	0.908	1.577	1.986	2.762	3.638
50	0.788	1.426	1.811	2.538	3.358	0.894	1.560	1.965	2.735	3.604

TABLA I (continuación) *k*-valores para los límites de tolerancia unilaterales cuando se muestrean distribuciones normales

$n \backslash d$	$\gamma = 0.95$ 0.75	0.90	0.95	0.99	0.999	$\gamma = 0.99$ 0.75	0.90	0.95	0.99	0.999
6	1.895	3.006	3.707	5.062	6.612	2.849	4.408	5.409	7.334	9.540
7	1.732	2.755	3.399	4.641	6.061	2.490	3.856	4.730	6.411	8.348
8	1.617	2.582	3.188	4.353	5.686	2.252	3.496	4.287	5.811	7.566
9	1.532	2.454	3.031	4.143	5.414	2.085	3.242	3.971	5.389	7.014
10	1.465	2.355	2.911	3.981	5.203	1.954	3.048	3.739	5.075	6.603
11	1.411	2.275	2.815	3.852	5.036	1.854	2.897	3.557	4.828	6.284
12	1.366	2.210	2.736	3.747	4.900	1.771	2.773	3.410	4.633	6.032
13	1.329	2.155	2.670	3.659	4.787	1.702	2.677	3.290	4.472	5.826
14	1.296	2.108	2.614	3.585	4.690	1.645	2.592	3.189	4.336	5.651
15	1.268	2.068	2.566	3.520	4.607	1.596	2.521	3.102	4.224	5.507
16	1.242	2.032	2.523	3.463	4.534	1.553	2.458	3.028	4.124	5.374
17	1.220	2.001	2.486	3.415	4.471	1.514	2.405	2.962	4.038	5.268
18	1.200	1.974	2.453	3.370	4.415	1.481	2.357	2.906	3.961	5.167
19	1.183	1.949	2.423	3.331	4.364	1.450	2.315	2.855	3.893	5.078
20	1.167	1.926	2.396	3.295	4.319	1.424	2.275	2.807	3.832	5.003
21	1.152	1.905	2.371	3.262	4.276	1.397	2.241	2.768	3.776	4.932
22	1.138	1.887	2.350	3.233	4.238	1.376	2.208	2.729	3.727	4.866
23	1.126	1.869	2.329	3.206	4.204	1.355	2.179	2.693	3.680	4.806
24	1.114	1.853	2.309	3.181	4.171	1.336	2.154	2.663	3.638	4.755
25	1.103	1.838	2.292	3.158	4.143	1.319	2.129	2.632	3.601	4.706
30	1.059	1.778	2.220	3.064	4.022	1.249	2.029	2.516	3.446	4.508
35	1.025	1.732	2.166	2.994	3.934	1.195	1.957	2.431	3.334	4.364
40	0.999	1.697	2.126	2.941	3.866	1.154	1.902	2.365	3.250	4.255
45	0.978	1.669	2.092	2.897	3.811	1.122	1.857	2.313	3.181	4.168
50	0.961	1.646	2.065	2.863	3.766	1.096	1.821	2.296	3.124	4.096

Source: G. J. Lieberman, *Table for one-sided statistical tolerance limits,* Industrial Quality Control XIV, 1958, 7–9. Reprinted with permission.

TABLA J Valores de cuantiles superiores de la distribución de la estadística D_n de Kolgomorov-Smirnov

	1 − α				
n	0.80	0.85	0.90	0.95	0.99
1	.900	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.828
4	.494	.525	.564	.624	.733
5	.446	.474	.510	.565	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.410	.490
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.361	.433
14	.274	.292	.314	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.392
17	.250	.266	.286	.318	.381
18	.244	.259	.278	.309	.371
19	.237	.252	.272	.301	.363
20	.231	.246	.264	.294	.356
25	.21	.22	.24	.27	.32
30	.19	.20	.22	.24	.29
35	.18	.19	.21	.23	.27
Fórmula para una *n* mayor	$\frac{1.07}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Fuente: F. J. Massey, Jr., *The Kolmogorov-Smornov test for goodness of fit,* J. Amer Statistical Assoc. 46 (1951), 68-78. Publicado con permiso.

TABLA K Límites de la estadística de Durbin-Watson

$1 - \alpha = 0.95$

	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

TABLA K (continuación) Límites de la estadística de Durbin-Watson

	$1 - \alpha = 0.99$									
	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
16	0.84	1.09	0.74	1.25	0.63	1.44	0.53	1.66	0.44	1.90
17	0.87	1.10	0.77	1.25	0.67	1.43	0.57	1.63	0.48	1.85
18	0.90	1.12	0.80	1.26	0.71	1.42	0.61	1.60	0.52	1.80
19	0.93	1.13	0.83	1.26	0.74	1.41	0.65	1.58	0.56	1.77
20	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
21	0.97	1.16	0.89	1.27	0.80	1.41	0.72	1.55	0.63	1.71
22	1.00	1.17	0.91	1.28	0.83	1.40	0.75	1.54	0.66	1.69
23	1.02	1.19	0.94	1.29	0.86	1.40	0.77	1.53	0.70	1.67
24	1.04	1.20	0.96	1.30	0.88	1.41	0.80	1.53	0.72	1.66
25	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
26	1.07	1.22	1.00	1.31	0.93	1.41	0.85	1.52	0.7[illegible]	1.64
27	1.09	1.23	1.02	1.32	0.95	1.41	0.88	1.51	0.81	1.63
28	1.10	1.24	1.04	1.32	0.97	1.41	0.90	1.51	0.83	1.62
29	1.12	1.25	1.05	1.33	0.99	1.42	0.92	1.51	0.85	1.61
30	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
31	1.15	1.27	1.08	1.34	1.02	1.42	0.96	1.51	0.90	1.60
32	1.16	1.28	1.10	1.35	1.04	1.43	0.98	1.51	0.92	1.60
33	1.17	1.29	1.11	1.36	1.05	1.43	1.00	1.51	0.94	1.59
34	1.18	1.30	1.13	1.36	1.07	1.43	1.01	1.51	0.95	1.59
35	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	0.97	1.59
36	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	0.99	1.59
37	1.22	1.32	1.16	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.23	1.33	1.18	1.39	1.12	1.45	1.07	1.52	1.02	1.58
39	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
45	1.29	1.38	1.24	1.42	1.20	1.48	1.16	1.53	1.11	1.58
50	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
55	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.41	1.47	1.38	1.50	1.35	1.53	1.31	1.57	1.28	1.61
70	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
75	1.45	1.50	1.42	1.53	1.39	1.56	1.37	1.59	1.34	1.62
80	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
85	1.48	1.53	1.46	1.55	1.43	1.58	1.41	1.60	1.39	1.63
90	1.50	1.54	1.47	1.56	1.45	1.59	1.43	1.61	1.41	1.64
95	1.51	1.55	1.49	1.57	1.47	1.60	1.45	1.62	1.42	1.64
100	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

Fuente: J. Durbin and G. S. Watson, *Testing for serial correlation in least squares regression, II,* Biometrika **38** (1951), 159-178. Publicado con permiso de Biometrika Trustees.

Respuestas a los ejercicios seleccionados de número impar

Capítulo 1

1.1. *a), b)*

Límites verdaderos	*Frecuencia de la clase*	*Frecuencia relativa*	*Frecuencia relativa acumulativa*
(0.15, 1.55)	17	0.34	0.34
(1.55, 2.95)	11	0.22	0.56
(2.95, 4.35)	7	0.14	0.70
(4.35, 5.75)	6	0.12	0.82
(5.75, 7.15)	4	0.08	0.90
(7.15, 8.55)	3	0.06	0.96
(8.55, 9.95)	2	0.04	1.00
Totales	50	1.00	

c) Los intervalos intercuantil e interdecil son, en forma aproximada, iguales a 3.8 min y 6.7 min, respectivamente.

d) $\bar{x} = 3.258$; *Mediana* $= 2.6182$; *Moda* $= 0.85$; $s = 2.4986$; *M.D.* $= 2.081$; y *Md.D.* $= 2.0042$

e) $\bar{x} = 3.26$; *Median* $= 2.75$; *Moda* $= 0.4$; $s = 2.4819$; *M.D.* $= 2.0056$; y *Md.D.* $= 1.948$

1.3. $\bar{x} = 3.5$; las varianzas son $s_1^2 = 3.5$, $s_2^2 = 7.5$, y $s_3^2 = 109.9$

1.5. *a)*

Límites verdaderos	*Frecuencia*	*Frecuencia relativa*
(−1.875, −1.125)	5	0.1667
(−1.125, −0.375)	5	0.1667
(−0.375, 0.375)	8	0.2667
(0.375, 1.125)	8	0.2667
(1.125, 1.875)	4	0.1333
Totales	30	1.0001

Ningún cambio en la distribución de frecuencia relativa

1.7. *b)* $\bar{x} = 18.82$; *Moda* = 14.0
c) $s^2 = 123.4196$; $s = 11.11$; *M.D.* = 9.27

Capítulo 2

2.1. *a)* Los eventos no son mutuamente excluyentes
b) 1. 180/400; 2. 150/400; 3. 30/400; 4. 60/180; 5. 60/200
c) $P(S \mid M) = 50/220$; $P(S) = 150/400$; no son estadísticamente independientes
d) No; $P(A \mid F) = 0.1111$, $P(A) = 0.125$
e) 1. 240/400; 2. 210/400; 3. 60/400; 4. 30/50

2.3. Cuando alguno o ambos eventos son vacíos

2.5. Las permutaciones son GGG, GGB, GBG, BGG, BBG, BGB, y BBB. La probabilidad de tener dos niños del mismo sexo es 6/8; la probabilidad de un niño y dos niños es 3/8; la probabilidad de que todos sean del mismo sexo es 2/8.

2.7. $(1/2)^{10}$; 1/2

2.9. 13/30

2.11. *a)* Cuatro resultados posibles: ambos componentes trabajan; ambos no y uno trabaja y el otro no (en dos formas posibles)
b) 0.99

2.13. $n = 4$

2.15. 0.6571

2.17. 0.41

Capítulo 3

3.1. *a), c)*

x	$p(x)$	$F(x)$
0	0.0498	0.0498
1	0.1494	0.1992
2	0.2240	0.4232
3	0.2240	0.6472
4	0.1680	0.8152
5	0.1008	0.9160
6	0.0504	0.9664
7	0.0216	0.9880

3.3. *a)* 3/2; *b)* $(x^3 + 1)/2$; *c)* 7/16, 1/8

3.5. *a)* $(1/100)\exp(-x/100)$; *b)* 0.8187

3.7. $E(X) = 4$; $Var(X) = 4.1$

3.9. *a)* 1/3; *b)* 1/18

3.11. *a)* 4; *b)* 16; *c)* 2; *d)* 9
e) La distribución del ejercicio 3.10 es simétrica y se encuentra centrada alrededor del valor 5, tiene varianza igual a 8.33 y desviación estándar de 2.8868. Esta distribución tiene un sesgo positivo y un pico relativamente grande; la dispersión relativa también es grande.

3.13. *a)* $\sigma^2 + (\mu - c)^2$; *b)* $c = \mu$

3.15. $M.D.(X) = 0.19753$, $d.e.\ (X) = 0.2357$

3.17. *a)* Media = 800, Mediana = 554.52; *b)* 878.89 *c)* 1757.78; *d)* 0.3679

3.19. *a)* $(1 - 4t)^{-2}$ *b)* $E(X) = 8$, $Var(X) = 32$

Capítulo 4

4.5. *a)* 0.6562, 0.9346; *b)* 0.5696, 0.9391

4.7. $P(X = 4) = 0.0049$, $P(X \geq 4) = 0.0055$; existe una inclinación a concluir que la afirmación es incorrecta

4.9. 0.2122

4.11. Seis o más

4.15. 0.7601, 0.9718

4.17. 0.0488

4.19. 0.0803, 0.9862

4.21. 0.6767

4.23. 0.0293, sí

4.25. 0.1837, no

4.27. *a)* 0.5973; *b)* 5987; *c)* 0.6065

4.31.

x	*Frecuencia relativa*	*Probabilidad teórica*
0	0.715	0.7201
1	0.179	0.1689
2	0.063	0.0630
3	0.019	0.0263
4	0.010	0.0116
5	0.010	0.0053
6	0.002	0.0025
7	0.000	0.0012
8	0.002	0.0006

4.33. *a)* 0.0189; *b)* 0.0180; la ocurrencia es poco probable

Capítulo 5

5.3. *a)* 0.4649; *b)* 0.2204; *c)* 0.0228; *d)* 0.8643

5.5. *a)* 1.775; *b)* 18.225; *c)* 21.65; *d)* −1.65; *e)* 0.2; *f)* 19.8

5.7. 1018

5.9. 0.00069

5.11. 0.000008; la ocurrencia es muy poco probable

5.13. $228 000

5.15. *a)* 0.0256; *b)* ≈0; *c)* ≈0

5.17. Sí, la probabilidad de ocurrencia es virtualmente cero.

5.19. *a)* 0.5774; *b)* no

5.21. *a)* = 4, *b* = 16

5.23. *b)* $E(X) = 0.75$, $Var(X) = 0.0375$, $D.M.(X) = 0.1582$, $\alpha_3(X) = -0.8607$, $\alpha_4(X) = 3.0952$
c) 0.6679, 0.9523; *d)* 0.63, 0.7937, 0.9086

5.25. 0.64, 0.9728

5.27. 0.1314, 0.0582

5.29. *a)* 0.594; *b)* 0.0466; *c)* 0.2642

5.31. $\alpha = 3.75$, $\theta = 8$

5.35. 0.9409

5.37. *a)* $E(X) = 44.3113$; 16.23, 23.62, 29.86, 35.74, 41.63, 47.86, 54.86, 63.43, 75.87
b) 0.1054

5.39. *a)* 0.3679; *b)* 0.8647, 0.9502

5.41. *a)* 0.1353; *b)* 433

5.45. Exponencial con parámetro de escala θ^{α}

Capítulo 6

6.1. 0.0022; la ocurrencia es poco probable

6.3. *a)* $F(x, y) = (3x^2y - xy^2 - 3x^2 + x - 3y + y^2 + 2)/10$
b) 0.225
c) $F_X(x) = (3x - 1)(x - 1)/5$, $F_Y(y) = (9y - y^2 - 8)/10$
d) $f_X(x) = (6x - 4)/5$, $f_Y(y) = (9 - 2y)/10$

6.5. *a)* $p_X(x) = p_Y(y) = 5/16, 6/16, 5/16$, $x = y = -1, 0, 1$, respectivamente
b) no; *c)* 0

6.7. *a)* 0.69; *b)* $f_{T_1}(t_1) = (1/5)\exp(-t_1/5)$, $f_{T_2}(t_2) = 10\exp(-10t_2)$

6.9. 1029.2152

6.13. Si $Cov(X, Y) > 0$, $Var(X + Y) > Var(X) + Var(Y)$ y $Var(X - Y) < Var(X) + Var(Y)$; si $Cov(X, Y) < 0$, $Var(X + Y) < Var(X) + Var(Y)$, y $Var(X - Y) > Var(X) + Var(Y)$

6.15. 11/27

6.17. *a)* $\mu = 0.04$, $\sigma^2 = 0.0014769$; *b)* $f(p \mid x) = 1260p(1 - p)^{34}$
c) $\mu = 0.054$, $\sigma^2 = 0.0013456$; *d)* 0.5432

6.19. *a)* 1/2; *b)* 50
c) $f(x \mid y) = \exp\{-(1/150)[x - 50 - (y - 25)/2]^2\}/\sqrt{150\pi}$
d) $f(x \mid Y = 30) = \exp[-(1/150)(x - 52.5)^2]/\sqrt{150\pi}$, 0.9251

Capítulo 7

7.3. *a)* $\lambda^{\Sigma x_i}\exp(-n\lambda)/\Pi x_i!$; *b)* $p^n(1 - p)^{\Sigma x_i}$
c) $1/(b - a)^n$; *d)* $(1/\sigma\sqrt{2\pi})^n \exp[-\Sigma(x_i - \mu)^2/2\sigma^2]$

7.5. Partes *c)*, *b)*, y *f)*

7.11. 0.0075

7.15. *a)* 0.7698; *b)* 0.9986; *c)* 0.0548; *d)* 0.0228

7.17. ≈0; el inspector debe emprender la acción apropiada

7.19. 255.82

7.23. 0.99

7.25. Muy poco probable $P(T > 3.429) < 0.005$

7.27. Si; $P(T < -3.516) < 0.001$

7.29. Dudoso; $P(F_{15,20} > 1.999) < 0.10$

Capítulo 8

8.1. *a)* $ECM(T_1) = p(1 - p)/n$; $ECM(T_2) = [np(1 - p) + (2p - 1)^2]/(n + 2)^2$
b) No. Para $n = 10$, si $0.138 < p < 0.862$, $ECM(T_2) < ECM(T_1)$; de otro modo $ECM(T_1) < ECM(T_2)$. Para $n = 25$, si $0.142 < p < 0.858$, $ECM(T_2) < ECM(T_1)$; de otro modo, $ECM(T_1) < ECM(T_2)$

8.5. T_3; $Var(T_3)/Var(T_1) = 0.9$

8.9. $\hat{\lambda} = \overline{X}$

8.11. $\hat{\sigma}^2 = \Sigma X_i^2/2n$; sí

8.13. Los factores de la muestra de forma son – 0.0028 y 2.21, respectivamente; la distribución es, es forma aparente, simétrica y ligeramente plana en su parte superior.

8.15. *a)* $\hat{\theta} = 100.0696$; *b)* sí, $\tilde{\theta} = 103.575$; *c)* 0.1057

8.17. *a)* 2532.7; *b)* 0.2061

8.19. *a)* 214.9289; *b)* 0.8410, 0.5340

8.21. (20.1191, 20.6434)

8.27. (151.31, 165.69), (149.75, 167.25), (147.82, 169.18)

8.31. (−3.89, −1.51), (−4.12, −1.28), (−4.58, −0.82), sí

8.33. (4.84, 21.16), (2.07, 23.93), sí

8.35. (146.98, 645.69)

8.39. (0.2048, 4.0744), sí

8.41. (0.0172, 0.0628), (0.0128, 0.0672), (0.0043, 0.0757); existe una razón para dudar det la afirmación

8.43. 663 8.45. *a)* 88; *b)* (66.40, 109.60)

8.47. (2.98514, 3.01486) 8.49. 0.8609, 299 8.51. 152

Capítulo 9

9.1. Prueba *b*

9.3. *a)* H_0: $p = 0.05$ contra H_1: $p > 0.05$; *b)* 0.2642
c) 0.3396, 0.4831, 0.6083, 0.8244, 0.9308, y 0.9757, respectivamente
d) $\alpha = 0.0755$, la potencia es 0.1150, 0.2120, 0.3231, 0.5951, 0.7939 y 0.9087, respectivamente.

9.5. Los valores críticos para la prueba 1 son 19.8333 y 20.1667; para la prueba 2 éstos son 19.7917 y 20.2083.

	Potencia										
μ	19.5	19.6	19.7	19.8	19.9	20.0	20.1	20.2	20.3	20.4	20.5
Prueba 1	≈1	0.9974	0.9452	0.6554	0.2126	0.0456	0.2126	0.6554	0.9452	0.9974	≈1
Test 2	0.9998	0.9893	0.8643	0.4602	0.0968	0.0124	0.0968	0.4602	0.8643	0.9893	0.9998

9.7. El extremo izquierdo de la distribución de muestreo de $\overline{X}$

9.9. *a)* H_0: $\lambda = 2.5$ contra H_1: $\lambda < 2.5$
b) Para las cuatro semanas, el valor crítico es 5
c) 0.8088

9.11. No puede rechazarse H_0, ya que $\bar{x} = 145 < \bar{x}_0 = 233.8$

9.13. 30

9.15. 7

9.17. 100.62; H_0 no puede rechazarse si el valor propuesto es $\geqslant$ 100.62

9.19. *a)* Valores relativamente grandes de manera que es fácil rechazar H_0
b) $t = 0.667$; H_0 no puede ser rechazada con $\alpha = 0.1$; el valor p es mayor que 0.2
c) Sí; los valores extremos pueden ser críticos

9.21. $t = 0.54$; H_0 no puede rechazarse

9.23. *a)* Sí; el valor crítico es tres, el valor p es 0.0755
b) $z = 2.05$ y H_0 se rechaza, el valor p es 0.0202

9.25. $z = -3.54$, H_0 se rechaza

9.27. $z = 1.62$, H_0 no puede rechazarse, el valor p es 0.1052

9.29. $[(z_{1-\beta} + z_{1-\alpha})^2(\sigma_X^2 + \sigma_Y^2)]/(\delta_0 - \delta_1)^2$

9.31. $t = -1.36$, H_0 no puede rechazarse; el valor p es 0.19

9.33. $t = -1.729$, H_0 no puede rechazarse

9.37. *a)* $t = 2.11$, H_0 se rechaza; *b)* el valor p es 0.039; *c)* (1.66, 7.84)

9.39. $\chi^2 = 17.28$, H_0 no puede rechazarse

9.41. Aproximadamente 0.1

9.43. *a)* $\chi^2 = 47.04$, H_0 se rechaza
b) los valores en el intervalo (1.8083, 7.1666); no son equidistantes debido a que la distribución de muestreo no es simétrica

9.45. *a)* $f = 3.24$, H_0 se rechaza; *b)* 0.1374

9.47. $z = 3.03$, H_0 se rechaza; el valor p es 0.0012

9.49. $z = 1.33$, H_0 no puede rechazarse; el valor p es 0.1836

Capítulo 10

10.1. $\chi^2 = 12$, H_0 se rechaza; el valor p es aproximadamente 0.008

10.3. *a)* $\chi^2 = 400$, H_0 se rechaza; la conclusión es diferente a la del ejercicio 10.2

10.5. $\chi^2 = 40$, H_0 se rechaza

10.7. $\chi^2 = 4.25$, H_0 no puede rechazarse

10.9. *a)* $\chi^2 = 5.8501$, H_0 no puede rechazarse
b) $\hat{\lambda} = 2.673$, $\chi^2 = 5.8969$, y H_0 a pesar de lo anterior, no puede rechazarse

10.11. La desviación máxima es 0.1263, H_0 no puede rechazarse

10.13. $\chi^2 = 1.0097$ (para $k = 5$ clases), H_0 no puede rechazarse

10.15. $\chi^2 = 7.8628$, H_0 no puede rechazarse

10.17.

2.16	3.78	2.70	1.36
2.60	4.54	3.24	1.62
3.24	5.68	4.06	2.02

10.19. $\chi^2 = 22.04$, H_0 se rechaza

10.21. $\chi^2 = 2.69$, H_0 no puede rechazarse

Capítulo 11

11.1. *a)* (4.3292, 5.6708), el promedio de la muestra de la decimoprimera semana es mayor que el valor del límite de control superior
b) probabilidad ≈ 0; c. 0.0301

11.3. *a)* (475.5051, 524.4949), no; *b)* 0.9884; *c)* (0.574, 37.486)

11.5. *a)* (378.36, 422.04), (0, 31.96); *b)* 16.28

11.7. *a)* (0, 0.0797); *b)* (0, 0.0758); *c)* 0.0013

11.9. *a)* 0.6767; *b)* 0.5438

11.11. 0.5526

11.13. Aproximadamente 0.175

11.15. $n = 99$, $c = 4$; $n = 131$, $c = 5$; $n = 100$, $c = 4$; $n = 116$, $c = 5$

11.17. *a)* 0.3679; *b)* 0.019; *c)* 0.3971; *d)* 0.216

11.19. $n = 65$, $\bar{x}_a = 71.53$

Capítulo 12

12.5.

Fuente	gl	SC	CM	*Valor F*
Tratamientos	2	0.492	0.246000	43.41
Error	12	0.068	0.005667	
Total	14	0.560	$f_{0.95,2,12} = 3.89$	

12.7. *a)*

Fuente	gl	SC	CM	*Valor F*
Tratamientos	3	2305.5	768.50	2.75
Error	28	7838.0	279.93	
Total	31	10143.5	$f_{0.95,3,28} = 2.95$	

b) Los residuos estandarizados no sugieren varianzas desiguales.

12.9. *a)*

Fuente	gl	SC	CM	*Valor F*
Tratamientos	4	522,744	130,686	66.41
Error	20	39,360	1,968	
Total	24	562,104	$f_{0.99,4,20} = 4.43$	

b) Algunos contrastes y sus intervalos de confianza son: $L_1 = \mu_5 - \mu_4$, (41.87, 278); $L_2 = 3\mu_5 - \mu_2 - \mu_3 - \mu_4$, (406.65, 985.35); $L_3 = \mu_2 - \mu_1$, (33.87, 270.13); $L_4 = 2\mu_5 - \mu_3 - \mu_4$, (205.39, 614.61); $L_5 = \mu_3 - \mu_2$, (−82.13, 154.13)

12.11. El uso del análisis de varianza es cuestionable debido a que la variación en el interior de la región es demasiado grande para ser atribuida sólo a un error aleatorio.

12.13. *b)* $Y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij}$, $i = 1, 2, \ldots, 5$, $j = 1, 2, 3, 4$; H_0: $\tau_j = 0$ para toda j;

Fuente	gl	SC	CM	*Valor F*
Bloques	4	1026.2875		
Tratamientos	3	17.6260	5.8753	41.09
Error	12	1.7165	0.1430	
Total	19	1045.6300	$f_{0.95,3,12} = 3.49$	

c) Algunos contrastes y sus intervalos de confianza son $L_1 = 3\mu_4 - \mu_1 - \mu_2 - \mu_3$, (4.48, 8.28); $L_2 = \mu_4 - \mu_1$, (1.57, 3.11); $L_3 = \mu_2 + \mu_3 - 2\mu_1$, (−0.70, 1.98); $L_4 = \mu_3 - \mu_2$, (−0.41, 1.13)

12.15. *a)* $Y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij}$, $i = 1, 2, \ldots, 7$, $j = 1, 2, 3, 4$

b)

Fuente	gl	SC	CM	*Valor F*
Bloques	6	1,471,772.429		
Tratamientos	3	44,826.572	14,942.19	16.48
Error	18	16,316.428	906.47	
Total	27	1,532,915.429	$f_{0.95,3,18} = 3.16$	

c) $f = 16.48 > f_{0.95,1,6} = 5.99$; H_0 a pesar de esto se rechaza

d) Dos contrastes y sus intervalos de confianza son: $L_1 = \mu_1 - \mu_3$, (12.01, 111.13); $L_2 = \mu_2 - \mu_4$, (−29.13, 69.99)

12.17. *a)* $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$, $i = 1, 2$, $j = 1, 2$, $k = 1, 2, 3$

b) H_0: $(\alpha\beta)_{ij} = 0$ para toda i y j; H_0: $\alpha_i = 0$ para toda i y j; H_0: $\beta_j = 0$ para toda j

c)

Fuente	gl	SC	CM	*Valor F*
Horno	1	0.022534	0.022534	3.92
Temperatura	1	0.005634	0.005634	0.98
Horno × Temp	1	0.554699	0.554699	96.47
Error	8	0.046000	0.005750	
Total	11	0.628867	$f_{0.95,1,8} = 5.32$	

12.19. *a)* $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$, $i = 1, 2, 3, 4$, $j = 1, 2, 3$, $k = 1, 2, 3, 4$

b) H_0: $\sigma^2_{\alpha\beta} = 0$; H_0: $\alpha_i = 0$ para toda i; H_0: $\sigma^2_\beta = 0$

c)

Fuente	gl	SC	CM	*Valor F*
Variedades	3	331.750	110.58	0.64
Fertilizantes	2	22,764.875	11,382.44	230.65
Variedades × Fert.	6	1,052.125	173.35	3.51
Error	36	1,776.500	49.35	
Total	47	25,925.250	$f_{0.95,3,6}$ = 4.76	

$f_{0.95,2,36} = 3.27$ $f_{0.95,6,36} = 2.37$

Capítulo 13

13.3. *a)* $\Sigma Y_i x_i / \Sigma x_i^2$; *b)* $E(B) = \beta$

13.5. *a)* En algún grado; *b)* $\hat{y} = 2.50 + 1.7774x$
c) Para cualquier aumento de $1 000 en el ingreso familiar, la cobertura del seguro de vida también aumenta.
d) Debe ajustarse una ecuación cuadrática

13.7. *a)*

x	45	20	40	40	47	30	25	20	15
residuos	−12.48	11.95	−13.60	−23.60	3.96	−0.82	8.06	−3.05	10.84

x	35	40	55	50	60	15	30	35	45
residuos	0.29	1.40	4.74	18.63	10.85	0.84	−15.82	0.29	−2.48

c) 124.7; *d)* 7.727, 0.2021; *e)* (1.3489, 2.2059); *f)* Sí, $t = 8.79$

g)

x_p	*Intervalo de confianza*	x_p	*Intervalo de confianza*
45	(75.67, 89.29)	35	(59.11, 70.31)
20	(29.23, 46.87)	40	(67.75, 79.45)
40	(67.75, 79.45)	55	(90.38, 110.14)
40	(67.75, 79.45)	50	(83.17, 99.57)
47	(78.73, 93.35)	60	(97.43, 120.87)
30	(49.69, 61.95)	15	(18.60, 39.72)
25	(39.65, 54.23)	30	(49.69, 61.95)
20	(29.23, 46.87)	35	(59.11, 70.31)
15	(18.60, 39.72)	45	(75.67, 89.29)

13.9.

x_p	$\hat{y}_{part}$	*intervalo de predicción del 95%*
18	34.49	(8.98, 60.00)
28	52.27	(27.71, 76.83)
38	70.04	(45.70, 94.38)
48	87.82	(62.95, 112.69)
58	105.59	(79.50, 131.68)

13.11. *b)* 0.2262; la asociación lineal es vaga

13.13. *a)* $(\sigma^2/n) \leq Var(\hat{Y}_p) \leq 2(\sigma^2/n)$;
b) $[(n+1)\sigma^2/n] \leq Var(\hat{Y}_{part}) \leq [(n+2)\sigma^2/n]$;
c) $b_0 = 15.5$, $b_1 = 5.1$

13.15. *b)* $\hat{y} = 12.75 + 19.875x$; *c)* sí, $t = 14.24$; *d)* (16.881, 22.869)

13.17. *a)* $r = -0.8829$

b)

Fuente	gl	SC	CM	*Valor F*
Regresión	1	5.64305	5.64305	63.65
Error	18	1.59595	0.08866	
Total	19	7.23900	$f_{0.99,1,18} = 8.29$	

13.19. *a)* $\hat{y} = -53.119 + 0.6639x$; *b)* sí

c) se detecta una autocorrelación positiva $d = 0.7075$

d) $\hat{y} = -45.116 + 0.6704x$

Capítulo 14

14.1. *a)* β_2 es no lineal

14.3. *a)* 3; b. -3.5

14.5. *a)*

Variable en el modelo	b_0	b_1	b_2
x_1	0.1619	0.1342	
x_2	0.6713		−0.0363
x_1, x_2	−0.1605	0.1487	0.0769

b) $f = 113.14$, rechazar H_0

c) $SCR(x_2 \mid x_1) = 0.0879, f = 14.63$; $SCR(x_1 \mid x_2) = 1.3366, f = 222.47$

d) $R^2 = 0.9496$

e) $\hat{y} = -0.1605 + 0.1487x_1 + 0.0769x_2$, $\hat{y}_p = \$518.85$, (\$462.7, \$575.0)

14.7. $$\begin{bmatrix} 15 & 42.00 & 55.0 \\ 42 & 188.08 & 140.8 \\ 55 & 140.80 & 219.0 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 8.070 \\ 32.063 \\ 28.960 \end{bmatrix}$$

14.9. *a)*

Variable en el modelo	R^2	SCE	CME	C_p
x_1	0.000	7326	407.02	114.07
x_2	0.229	5648	313.77	84.27
x_3	0.784	1581	87.82	12.06
x_4	0.748	1846	102.53	16.76
x_1, x_2	0.230	5641	332.00	86.20
x_1, x_3	0.802	1451	85.34	11.75
x_1, x_4	0.754	1800	106.00	17.97
x_2, x_3	0.785	1576	92.73	13.99
x_2, x_4	0.774	1653	97.20	15.36
x_3, x_4	0.869	958	56.34	3.00
x_1, x_2, x_3	0.802	1451	90.70	13.77
x_1, x_2, x_4	0.778	1624	102.00	16.85
x_1, x_3, x_4	0.885	846	52.88	3.02
x_2, x_3, x_4	0.870	950	59.38	4.87
x_1, x_2, x_3, x_4	0.885	845	56.33	5.00

c) $\hat{y} = -114.988 + 1.2657x_3 + 0.8414x_4$, $\hat{y}_{part} = 100.35$, (82.05, 118.64)

14.11. 0.0654, 0.0952

14.13. La gráfica revela una tendencia cuadrática $\hat{y} = 8.238 + 0.3126x - 0.001823x^2$; $\hat{y} = -27.55\%$, lo cual es absurdo

14.15. *a)* $\hat{y} = 5284.28 - 114.85x_1 - 78.67x_2 + 0.129x_3 + 0.189x_1^2 + 0.201x_2^2 + 2.63 \times 10^{-7}x_3^2 + 1.268x_1x_2 - 0.0017x_1x_3 - 0.0008x_2x_3$

b) La elección para la mejor ecuación se encuentra entre las dos siguientes:
$\hat{y} = 2163.98 - 56.47x_1 - 26.17x_2 + 0.0162x_3 + 0.6952x_1x_2 - 0.0005x_1x_3$
$\hat{y} = 1676.19 - 43.57x_1 - 19.77x_2 + 0.526x_1x_2 - 5.91 \times 10^{-5}x_1x_3$

14.17.

	Y	x_1	x_2	x_3	x_4	x_5
Y	1.00	0.18	0 78	0.15	−0.29	0.45
x_1	0.18	1.00	−0.05	0.41	0.55	−0.04
x_2	0.78	−0.05	1.00	0.08	−0.30	0.16
x_3	0.15	0.41	0.08	1.00	0.27	−0.14
x_4	−0.29	0.55	−0.30	0.27	1.00	−0.11
x_5	0.45	−0.04	0.16	−0.14	−0.11	1.00

Capítulo 15

15.1. Sí, $t = -2.12$ y se rechaza H_0

15.3. $z = -2.51$, H_0 se rechaza

15.5. $z = -2.80$ y, por lo tanto, existe una razón para creer que la secuencia no es aleatoria.

15.7. $z = 2.24$ y la H_0 de que no existe diferencia entre la preferencia se rechaza. La conclusión es diferente a la del ejercicio 15.6.

15.9. $s = 4$, los valores críticos son 2 y 10, no puede rechazarse H_0

15.11. $h = 11.40$ y se rechaza la hipótesis H_0 de que las distribuciones son idénticas para ambas disciplinas.

15.13. $s = 1.1$ y las diferencias entre las cuatro pruebas no son estadísticamente significativas.

15.15. $r_S = 0.7915$

15.17. $r_S = -0.4667$; existe alguna tendencia en uno de los jueces para dar un puntaje alto cuando los demás lo dan bajo.

Índice analítico